Subjective Financial Literacy in Housing Markets*

Boaz Abramson
Columbia Business School

Andres Yany
Amazon

June 2024

Abstract

This paper studies the role of subjective financial literacy in housing choices. We find that individuals who self-assess themselves as more financially literate are more likely to become homeowners and tend to take on more levered positions to finance their home acquisition. We solve a heterogeneous agent portfolio choice model to infer the mechanisms that underlie the empirical patterns. We find that households who self-assess themselves as more financially literate are in fact more financially savvy - they access cheaper and larger credit and earn higher risk-adjusted returns on their housing investments. We show that heterogeneity in financial literacy is important for reliably evaluating housing policies. The elasticity of homeownership with respect to wealth is roughly halved when heterogeneity in financial literacy is modeled.

JEL-CODES: E21,G50,G53,R21

Keywords: financial literacy, household heterogeneity, housing choice, expectations, credit constraints

*Corresponding author: Boaz Abramson, 665 West 130 Street, New York, NY 10027, email: ba2695@gsb.columbia.edu. For valuable guidance we thank Monika Piazzesi and Martin Schneider. We have also benefited from helpful suggestions by Adrien Auclert, Liran Einav, Pete Klenow, Sean Myers, Alessandra Peter, Chris Tonetti, Stijn Van Nieuwerburgh, as well as participants in various workshops and seminars.

1 Introduction

Buying a home and taking a mortgage are two of the most important financial decisions households make throughout their life. Home equity is the single largest asset on U.S. households' balance sheets, and mortgages are the single largest liability (Survey of Consumer Finances, 2022). Understanding the determinants of homeownership and mortgage choice is therefore of first order importance. Indeed, a voluminous literature has examined, for example, the role of credit and liquidity constraints (Campbell and Cocco, 2003; Yao and Zhang, 2005; Ortalo-Magne and Rady, 2006; Greenwald and Guren, 2024), age, income and wealth (Cocco, 2004; Adelino, Schoar and Severino, 2016), mobility (Stanton and Wallace, 1998), expectations (Glaeser and Nathanson, 2017; Bailey et al., 2019; Kuchler and Zafar, 2019; Gargano et al., 2023; Kuchler et al., 2023), peer effects (Bailey et al., 2018) and race and ethnicity (Charles and Hurst, 2002; Fuster et al., 2022; Bartlett et al., 2022) in explaining the observed variation in ownership and mortgage choices across households.

This paper studies the role of *subjective* financial literacy in individuals' housing choices. Subjective financial literacy refers to an individual's *belief* regarding her financial literacy. Exploiting a novel feature of the 2016 Survey of Consumer Finances (SCF), we document that subjective financial literacy is an important determinant of housing choices. Specifically, we find that individuals who believe that they are more financially literate are 1) more likely to own a house rather than rent one, and 2) tend to take on more levered positions to finance their home acquisition. The relationship is economically meaningful and statistically robust to controlling for a host of potential confounding factors.

We then ask what are the mechanisms through which subjective financial literacy impacts housing choices. We consider two main candidates. First, households that differ in their subjective financial literacy might have access to different mortgage terms. This would be the case, for example, if households who believe they are more financially savvy are in fact better at searching and negotiating for advantageous mortgage terms. Second, households with different levels of subjective financial literacy might hold different expectations on future house prices. This could reflect either differences in households' beliefs or differences in their true ability to search for better investment opportunities.

To examine the role of the different mechanisms in explaining the observed variation in housing choices, we solve a standard life-cycle dynamic-stochastic model of portfolio choice with housing (Campbell and Cocco, 2003; Cocco, 2004; Yao and Zhang, 2005). Households can consume housing services by owning or renting. One the one hand, buying a house requires incurring a larger upfront cost. Households can borrow to finance their home acquisition, but borrowing is subject to a collateral constraint and can be expensive. On the other hand, owning allows households to capitalize on house price appreciation while renters can only save in a risk-free asset. The decision whether to rent or own depends on households' resources, age, the mortgage terms they are offered, and their expectations on future house prices.

The key new feature in the model is that we incorporate heterogeneity in subjective financial literacy. Specifically, mortgage spreads and collateral constraints are allowed to depend on households' subjective financial literacy. Expectations regarding future house prices, namely the expected mean and variance of the idiosyncratic shock to future house prices, can also depend on subjective literacy. We assume that subjective financial literacy is innate and focus on how it impacts housing choices. An important limitation of our cross-sectional data is that we cannot observe whether, and how, financial literacy dynamically evolves.

We quantify the model using SCF micro data on balance sheets, income, and demographic characteristics of a representative sample of U.S. households. W categorize households into three groups based their subjective financial literacy: low, intermediate and high. We estimate four groups of parameters: 1) the expected mean of the idiosyncratic shock to house price growth, 2) the expected volatility of this shock, 3) the minimum collateral requirement, and 4) the mortgage spread. Each of these parameters can depend on the household's subjective financial literacy. We estimate the parameters using a Simulated Method of Moments design. The data moments we target are homeownership rates and loan-to-value (LTV) ratios across the three groups of subjective financial literacy and across the life cycle.

The model closely accounts for the empirical relationship between subjective financial literacy and housing choices. As in the data, households with higher subjective financial literacy are more likely to be homeowners and they take on larger mortgages relative to the value of their house. As in the data, the relationship holds true after controlling for income, wealth and age. We further validate our model by showing that, despite only targeting the unconditional correlation between subjective financial literacy and housing choices, the model also matches well the correlation conditional on income, wealth and age. While there might be additional dimensions of heterogeneity that can rationalize the data, we focus on the two that are arguably most intuitive - mortgage terms and house price expectations. Our results suggest that heterogeneity along these dimensions is able to account for the empirical relationship.

We find that households with higher subjective financial literacy face more attractive mortgage terms - the estimated mortgage spread and minimum collateral constraint are decreasing with subjective literacy. Households with high subjective financial literacy also have more optimistic expectations on future house price growth relative to households with low subjective financial literacy - they expect the idiosyncratic shock to house price growth to be drawn from a distribution with a higher mean and lower standard deviation. In terms of identification, differences in borrowing conditions are mostly identified by differences in homeownership rates in the data. Intuitively, the stringency of the collateral constraint governs to degree to which young, resource-constrained, households can access the owner-occupied market. The mortgage spread matters relatively more for the tenure decision of middle-aged and older households. These households expect their income to decline, would therefore like to save, but many of them still have outstanding mortgages on their homes. The extent to which they are willing to continue paying off their mortgage, instead of selling and becoming renters, largely depends on the cost of debt. Differences in expectations on future house prices are mostly identified from the cross-sectional variation in loan-to-value ratios. For households who choose to own, the return they expect on the housing asset governs how much they choose to lever. The expected volatility of the shock to house prices matters relatively more for the leverage decision of older households, for whom the net present value of the non-risky component of income is lower, while the expected mean of the shock matters relatively more leverage choices of young households.

Which of the two mechanisms is more important for explaining the empirical relationship between subjective financial literacy and housing choices? To answer this question, we consider two variants of our model. In the first, we shut off heterogeneity in expectations and continue to allow heterogeneity in borrowing conditions. In the second, we consider the analog case where only heterogeneity in expectations is allowed. When heterogeneity in expectations is shut off, the model's fit to the relationship between subjective literacy and housing choices is dampened for older households, but not for younger households. In contrast, when heterogeneity in borrowing conditions is ignored, the model's fit to the data deteriorates more for younger households relative to middle-aged and older households. We conclude that heterogeneity in expectations matters relatively more for explaining the link between subjective literacy and housing choices among middle-aged and older households, while heterogeneity in borrowing conditions is more important for accounting for the cross-sectional variation among younger households. Intuitively, borrowing conditions matter more for housing choices at the beginning of life, when households tend to borrow, and expectations on house price appreciation matter more later in life, when households are more prone to save.

Our analysis suggests that households that have a higher subjective financial literacy expect a better risk-adjusted return on housing investments. An important question is whether these expectations reflect over-optimism or rather households' true ability to search for better investment opportunities. In other words, does subjective financial literacy proxy distorted beliefs, or rather true savviness in housing investments. In the baseline model, we assume the later. That is, differences in expectations are aligned with the true distributions from which future house prices are drawn. To test whether distorted beliefs can explain the empirical patterns, we consider an alternative model where we allow for heterogeneous expectations but in which shocks to future house prices are drawn from a distribution that does not depend on self-assessed literacy. We find that the fit of the alternative model with respect to the data deteriorates. This suggests that subjective financial literacy proxies, at least to some extent, objective financial literacy.

We argue that accounting for heterogeneity in financial literacy is crucial for reliably evaluating housing policies. To see this, we compare our model to a benchmark portfolio choice model with housing where heterogeneity in financial literacy is ignored. We then compute the impact of counterfactual housing policies in both models. This exercise allows us to quantify the bias in policy evaluation that arises if we abstract from heterogeneity in financial literacy. The particular policy we focus on is a shock to households' wealth. The wealth shock proxies policies that are designed to encourage homeownership, for example income transfers, tax deductions or capital gains exemptions.

We find that the impact of a wealth shock on homeownership is downsized by approximately 40% when we incorporate heterogeneity in financial literacy. The reason is that, absent heterogeneity in financial literacy, the model substantially over-estimates the correlations between housing choices and wealth, income, and age relative to the data. This in turn leads to a biased evaluation of the impact of a wealth shock. The heterogeneous agent model, in contrast, is able to substantially reduce the bias in the correlation between ownership and wealth, income, and age relative to the data. As a result, it produces more reliable policy evaluations. More broadly, our results highlight that documenting heterogeneity in financial decision making and incorporating this heterogeneity into structural models is important for understanding the impact of economic policies (Gomes, Haliassos and Ramadorai, 2021).

Related Literature

Our paper relates to a large literature in household finance that studies the role of financial literacy in household decision making.[1] In a seminal paper, Lusardi and Mitchell

[1] See Lusardi and Mitchell (2014) for a review.

(2007) document that financial literacy can explain observed differences in retirement savings across households. Financial literacy has also been linked to a host of other favorable financial outcomes such as stock market participation (Calvet, Campbell and Sodini, 2007; Van Rooij, Lusardi and Alessie, 2011), wealth accumulation (Van Rooij, Lusardi and Alessie, 2012; Jappelli and Padula, 2013), portfolio diversification (Guiso and Jappelli, 2008; Gaudecker, 2015), and avoidance of common investment mistakes (Calvet, Campbell and Sodini, 2009). Financial literacy also plays an important role in housing markets. Households that are more financially literate are more likely to optimally decide whether to buy points (Agarwal, Ben-David and Yao, 2017) and whether to take a fixed-rate or an adjustable-rate mortgage (Guiso et al., 2022). Financial literacy is also positively associated with optimal refinancing behavior (Keys, Pope and Pope, 2016) and with homeownership (Gathergood and Weber, 2017).

Our contribution to this literature is twofold. First, we develop the first theory of financial literacy in the housing markets. We build on the standard life-cycle dynamic-stochastic model of portfolio choice with housing (Campbell and Cocco, 2003; Cocco, 2004; Yao and Zhang, 2005) and augment it with heterogeneity in financial literacy. The advantage of our structural approach is that it allows us to examine the mechanisms that underlie the empirical relationship between financial literacy and housing choices. Moreover, using the model, we are able to quantify the importance of heterogeneity in financial literacy for the evaluation of housing market policies. Delavande, Rohwedder and Willis (2008), Jappelli and Padula (2013) and Lusardi, Michaud and Mitchell (2017) develop theoretical models of financial literacy and portfolio choice, but abstract from housing.

Our second contribution is to evaluate the role of *subjective* financial literacy in housing choices. The literature typically uses *objective* measures of literacy. Such measures include, for example, the ability to compute compound interest, comprehend percentages, distinguish between nominal and real interest rates, and perceive the benefits of diversification (Lusardi and Mitchell, 2007; Hastings, Madrian and Skimmyhorn, 2013). In contrast to these test-based measures, we focus on subjective financial literacy. Subjective literacy refers to individuals' *beliefs* regarding their financial literacy. Subjective financial literacy has been shown to predict credit card usage (Allgood and Walstad, 2013), retirement savings (Parker et al., 2012), and portfolio choice (Van Rooij, Lusardi and Alessie, 2011; Allgood and Walstad, 2013). We find that subjective literacy is a robust predictor of housing choices. Importantly, we show that subjective financial literacy is more predictive of housing choices relative to the traditional objective literacy measures.

Subjective financial literacy can be more informative of housing choices relative to objective measures of literacy for two main reasons. First, individuals' subjective assessment of their financial literacy might be biased relative to their true literacy level, for example due to over-optimism (Agnew and Szykman, 2005). According to this reasoning, households make housing choices based on their biased beliefs regarding their financial literacy, rather than based on their true financial literacy. Alternatively, individuals might actually be better at assessing their true financial literacy relative to what an econometrician can assess using a limited set of questions. According to this explanation, households make housing choices based on their true financial literacy, and the traditional objective measures are biased estimates of true financial literacy. Our structural model allows distinguishing between the two explanations. We find that subjective financial literacy largely proxies true financial literacy. These results suggest that, relative to traditional test-based measures, asking individuals to self-assess their own financial literacy might be a more accurate method to measure their true financial literacy.

Finally, our paper relates to the literature on housing market expectations (see Kuchler, Piazzesi and Stroebel (2023) for a review). A main strand of this literature focuses on uncovering the determinants of housing market expectations. Recent house price developments (Case and Shiller, 1988; Armona, Fuster and Zafar, 2019), personal experience (Kuchler and Zafar, 2019), social interactions (Shiller, 2007; Bailey et al., 2018) and ownership status (Kindermann et al., 2021) have been linked to individuals' housing market expectations. We contribute to this literature by documenting a link between financial literacy and housing market expectations. Using our structural model, we infer that households that self-assess themselves as more financially literate hold more optimistic expectations on house price growth. Our findings suggest that housing market expectations, which are unobservable (Kuchler, Piazzesi and Stroebel, 2023), can be elicited based on individuals' self-assessed financial literacy.

The paper proceeds as follows. Section 2 presents stylized facts relating subjective financial literacy to individuals' housing choices. Section 3 introduces a heterogeneous agent life-cycle model of optimal portfolio choice with housing that can rationalize these facts. Section 4 discusses the model estimation. Section 5 uses the quantified model to study the mechanisms through which subjective financial literacy impacts housing choices. Section 6 evaluates the importance of incorporating heterogeneity in financial literacy for policy evaluation. Section 7 concludes.

2 Facts

We begin by analyzing the relationship between subjective financial literacy, homeownership, and mortgage choices. The 2016 SCF wave offers a novel approach to measuring subjective financial literacy and relating it to housing choices. The 2016 wave asks respondents the following:

> *"On a scale from zero to ten, where zero is not at all knowledgeable about personal finance and ten is very knowledgeable about personal finance, what number would you (and your partner) be on the scale?"*

Figure 1 plots the SCF data on subjective financial literacy and housing market outcomes. The proportion of households who own a house is illustrated in the left panel and the ratio of collateralized debt to house value for home owners is plotted on the right panel. The basic stylized fact is that households who self-assess themselves as more financially literate are 1) more likely to own a house and 2) tend to take a more levered position on their house.

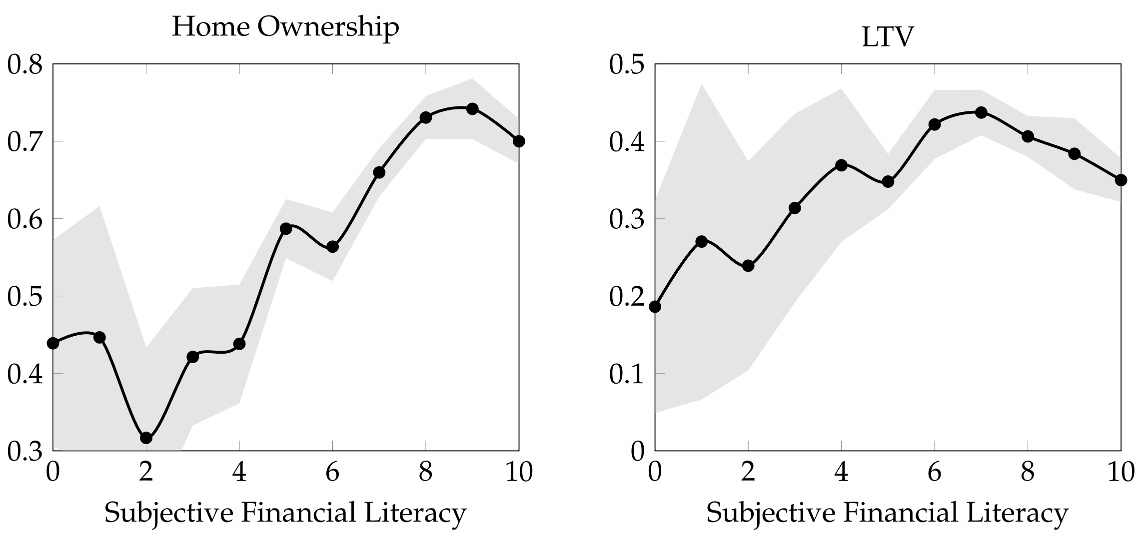

Figure 1: Subjective Financial Literacy in the Housing Markets

Notes: SCF data. Each dot represents the average homeownership rate (left panel) or loan-to-value ratio (right panel) conditional on self-assessed (subjective) financial knowledge. Home-ownership measures whether or not the household owns a ranch/farm/mobile home/house/condo. The Loan-To-Value ratio is computed for home owners as the ratio of housing collateralized debt to house value. Lines are kernel-weighted local polynomial regressions. 95% confidence intervals are plotted. Standard errors are computed using the "scfcombo" Stata package in order to account for the SCF complex sample specification as well as the multiple imputation process.

Table 1 provides descriptive statistics on the relationship between self-assessed financial literacy and additional socioeconomic characteristics. For ease of representation, households are classified into one of three groups according to whether they self-assess their financial literacy to be low (0-4 on scale, denoted by "Low FL"), intermediate (5-7, "Intermediate FL") or high (8-10, "High FL").[2] Households that self-assess themselves to

[2] While our results are robust to the exact pooling of households into groups, the data suggests a signifi-

be more financially literate are more educated. They also score higher in finance related questions that are often used to measure objective financial literacy.[3] This result is in line with previous work that finds a positive correlation between subjective and objective measures of financial literacy (Lusardi and Mitchell, 2011; Parker et al., 2012). Households with higher levels of self-assessed financial literacy report that they are willing to take on more risk, are more likely to use financial advisories, tend to participate more in the stock markets, are more likely to be males, and are wealthier. Unfortunately, the SCF does not collect information regarding expectations. The structural model we develop will allow us to infer whether households that differ in their subjective financial knowledge also differ in their expectations on future house prices and whether these expectation reflect fundamental differences in investment opportunities or heterogeneous beliefs.

One might worry that our subjective literacy measure is simply a linear combination of other, already observed, household characteristics. To alleviate these concerns, we regress self-assessed financial literacy on all the other variables reported in Table 1. The R^2 from this regression is only 0.1. This suggests that there is substantial cross sectional variation in subjective financial literacy that is unexplained by these observables.

We now argue that subjective financial literacy is an important and robust predictor of homeownership and mortgage choices. That is, the relationship illustrated in Figure 1 is not due to potential confounders. To establish this argument, we examine the linear cross-sectional relationships between self-assessed financial literacy and these housing market outcomes. We specify the following linear model:

$$Y_i = \beta_{low} FK_{low,i} + \beta_{high} FK_{high,i} + \Gamma X_i + \epsilon_i, \tag{1}$$

where Y_i is the outcome of interest, $FK_{low,i}$ is an indicator equal to one in case the household reports its financial literacy to be low (0-4 on the 0-10 scale), and $FK_{high,i}$ is the equivalent for households that self-assess their financial literacy to be high (8-10 on the scale). The omitted group consists of the intermediate literacy types. The vector of covariates X_i consists of an age polynomial, education attainment levels, a gender dummy,

cant intra-group variation in outcomes, larger than the inter-group variability.

[3] We define the financial literacy score as the number of correct answers to the following questions: 1) Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than today, exactly the same as today, or less than today with the money in this account? 2) Suppose you had \$100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than \$102, exactly \$102, or less than \$102? 3) Do you think that the following statement is true or false: buying a single company's stock usually provides a safer return than a stock mutual fund?

Table 1: Descriptive Statistics

Dependent Variable	Low FL	Intermediate FL	High FL
A. Demographics			
Age	51.80 (16.8)	50.77 (16.2)	54.47 (16.4)
Gender	0.62 (0.48)	0.71 (0.45)	0.74 (0.44)
Income	48,867 (132,113)	68,203 (64,066)	78,564 (79,444)
Wealth (log)	10.00 (2.64)	11.15 (2.11)	11.72 (2.00)
Education Level	2.26 (1.07)	2.79 (1.03)	2.87 (1.02)
B. Financial Indicators			
Objective Financial Literacy Score	1.86 (0.88)	2.12 (0.87)	2.24 (0.84)
Self-Assessed Financial Risk	2.79 (2.66)	4.09 (2.53)	4.37 (2.87)
Use of Advisories: Borrowing	0.40 (0.49)	0.52 (0.50)	0.59 (0.49)
Use of Advisories: Investing	0.44 (0.50)	0.58 (0.49)	0.63 (0.48)
Stock Market Participation	0.28 (0.49)	0.52 (0.50)	0.55 (0.49)
Equity Share of Financial Assets	0.44 (0.31)	0.43 (0.29)	0.43 (0.28)
Number of Stocks Held	0.48 (1.90)	0.78 (3.43)	1.58 (6.22)
Number of Observations	2,168	10,083	12,136

Notes: Households are classified into three groups according to their subjective financial literacy: Low (0-4 on scale), intermediate (5-7) and high (8-10). Income is the sum of wage income, income from retirement and social security funds, from self managed businesses and transfers. The "Gender" row reports proportion of males. Total wealth is defined by the SCF as the balance between total assets and total debt. The education level is a categorical variable that ranges from 1 (no high-school) to 4 (academic degree). The objective financial literacy score is measured as the number of correct answers to the three questions specified in footnote 3. Self-assessed financial risk is reported by households on a 0-10 scale, where 0 is "not at all willing to take financial risk". Use of financial advisories is a dummy equal one if the household reports using advisers when borrowing/investing. Stock market participation is an indicator equal to one if the household has equity in directly held stocks or mutual funds. Equity share is the ratio of equity to total financial assets. Capital gains are the nominal dollar gains on directly held stocks and mutual funds. Number of stocks measures the number of different directly held stocks in a household's portfolio.

total wealth and income, self-assessed risk preference, the traditional measures of objective financial literacy and dummies for usage of financial advisers when investing and borrowing. ϵ_i is the normally distributed error term.[4]

Table 2 reports the main results. Consistent with Figure 1, the first (second) column shows that the unconditional correlation between financial literacy and home ownership (LTV) is positive. As illustrated in the figure, differences in ownership rates are more stark than differences in loan-to-value ratios. Households who self-report high levels of financial literacy are more likely to own a house relative to those in the intermediate range (the change in odds ratio is 1.64), which are in turn more probable to be owners relative to those self-selecting to the low category (by an estimated change in odds ratio of 2.24). In terms of LTV, there doesn't seem to be much difference between the high and intermediary literacy types. Conditional on owning a house, the loan-to-value ratio of low types is 7.9% lower than that of the benchmark intermediate group.

Columns 3-4 then add the demographic controls, as well as education attainment levels.[5] Indeed, the magnitude of the relationship between self-assessed financial literacy and housing outcomes is weakened. However, the coefficients β_{low} and β_{high} are still economically and statistically significant. To interpret the sizable coefficients, the intermediate literacy households are 48% more likely to own a house with respect to low literacy types and are 24% less likely be home owners relative to the high types. This suggests that financial literacy is a source of heterogeneity that is economically meaningful for ownership and leverage over and above its relation to age, wealth, income, education and gender.

In columns 5-6, we also control for the usage of financial intermediaries and for households' objective financial literacy. If self-assessed literacy and the traditional objective measures of literacy are equivalent measures of sophistication, we should expect β_{low} and β_{high} to converge to zero. Not only is this not the case, but rather the objective measures are not as powerful in predicting home ownership as the self-assessed measure. The way people *self-assess* their financial literacy matters more for housing market outcomes. Finally, in columns 7-8 we also control for households willingness to take risk. The results show that self-assessed financial literacy does not simply proxy risk preferences.[6]

Taken together, the results presented in this section suggest that self-assessed financial literacy is an economically important and robust predictor of homeownership and

[4] In order to account for both the multiple imputation process and the dual-frame complex sample which are features of the SCF data, standard errors are computed using the "scfcombo" Stata package.

[5] We also control for wealth quartiles and an age polynomial.

[6] To further alleviate such concerns, our results are robust to additionally controlling for participation in the stock market and for equity shares.

Table 2: Prediction Regressions: Subjective Financial Literacy in the Housing Markets

	(1) Ownership	(2) LTV	(3) Ownership	(4) LTV	(5) Ownership	(6) LTV	(7) Ownership	(9) LTV
Subjective Fin. Lit.								
Low	-0.805^{***} (0.085)	-0.079^{**} (0.316)	-0.392^{***} (0.142)	-0.051^{***} (0.022)	-0.402^{***} (0.143)	-0.047^{***} (0.22)	-0.446^{***} (0.143)	-0.048^{***} (0.21)
High	0.494^{***} (0.051)	-0.025^{**} (0.011)	0.217^{**} (0.088)	0.011 (0.009)	0.208^{**} (0.090)	0.008 (0.009)	0.220^{**} (0.090)	0.005 (0.009)
Educ. Level								
High-School			-0.049 (0.130)	0.037^{**} (0.016)	-0.074 (0.132)	0.032^{**} (0.017)	-0.084 (0.132)	0.024^{**} (0.014)
Some College			-0.215 (0.131)	0.064^{***} (0.017)	-0.229^{*} (0.131)	0.056^{***} (0.017)	-0.220^{*} (0.131)	0.037^{**} (0.015)
Bachelors+			-0.619^{***} (0.158)	0.103^{***} (0.019)	-0.655^{***} (0.164)	0.090^{***} (0.019)	-0.623^{***} (0.165)	0.052^{*} (0.016)
Age			0.043^{***} (0.011)	-0.001 (0.001)	0.042^{***} (0.011)	-0.002 (0.002)	0.043^{***} (0.011)	-0.01^{***} (0.002)
Male			0.146 (0.101)	0.000 (0.011)	0.130 (0.102)	-0.004 (0.011)	0.155 (0.104)	-0.012 (0.01)
$ln(wealth)$			1.209^{***} (0.117)	-0.105^{***} (0.015)	1.210^{***} (0.118)	-0.109^{***} (0.015)	1.220^{***} (0.118)	-0.0104 (0.013)
$ln(income)$			-0.119^{*} (0.069)	0.160^{***} (0.006)	-0.122^{*} (0.071)	0.157^{***} (0.006)	-0.121^{*} (0.071)	0.119 (0.006)
Objective Fin. Lit.								
Inflation					0.218 (0.182)	-0.042^{*} (0.023)	0.207 (0.184)	-0.045^{**} (0.022)
Interest Rate					0.124 (0.173)	-0.073^{***} (0.024)	0.121 (0.175)	-0.066^{***} (0.023)
Diversification					0.116 (0.198)	-0.062^{**} (0.026)	0.123 (0.202)	-0.073^{***} (0.026)
Ad. Borrowing					0.318^{***} (0.079)	0.027^{***} (0.010)	0.328^{***} (0.079)	0.008 (0.008)
Ad. Investing					-0.222^{***} (0.073)	-0.002 (0.008)	-0.205^{***} (0.073)	0.007 (0.008)
Self. Ass. Fin. Risk							-0.051^{***} (0.013)	0.007^{***} (0.002)
Observations	24,112	15,007	24,112	15,007	24,112	15,007	24,112	15,007
R^2	0.055	0.048	0.414	0.384	0.417	0.389	0.418	0.489

Notes: Ownership measures whether or not the household owns a ranch/farm/mobile home/house/condo. The Loan-To-Value ratio is computed for home owners as the ratio of housing collateralized debt to house value. *** is significant at 1%; ** is significant at 5%; * is significant at the 10% level. Standard errors are computed using the "scfcombo" Stata package in order to account for the SCF complex sample specification as well as the multiple imputation process. The explanatory variables are subjective financial literacy (low, high), education level (high school, some college, bachelors), age, gender, ln(wealth), ln(income), objective financial literacy questions (inflation, interest rate, diversification), use of advisories (borrowing, investing) and self-assessed financial risk.

mortgage choices.[7] Households that self-assess themselves as more financially literate are more likely to own a house, and take on larger mortgages. We note that we are by no means claiming to identify a causal relationship between self-assessed financial literacy and housing outcomes.

3 Model

Motivated by the empirical patterns, we now ask what are the channels through which subjective financial literacy matters for housing tenure and leverage choices. We consider two intuitive candidates. First, households reporting higher levels of financial literacy might search for better mortgage terms, thereby paying lower interest rates on their mortgages and facing laxer collateral requirements. Second, expectations on future house values might be important. Households with different subjective financial literacy might expect different risk-return trade-offs in the housing markets, either due to access to different types of investment opportunities or due to distorted beliefs. While there might be other dimensions of household heterogeneity that can rationalize our empirical findings, we focus on those which are arguably most intuitive - mortgage terms and house price expectations. To examine the role of these different mechanisms in explaining the observed variation in housing choices, we solve a standard heterogeneous life-cycle model of portfolio choice with housing. The key novel feature of the model is that we introduce heterogeneity in subjective financial literacy.

3.1 Household Problem

Households live for a finite number of periods A. Time is discrete and indexed by t. Household age at time t is denoted by a_t. The probability of survival from period $t-1$ to period t is λ_{a_t}, and $\lambda_{a_{A+1}} = 0$. Household i enters the model with an innate level of subjective financial literacy f_i. Our model abstracts from the possibility that financial literacy might evolve throughout life. While literacy is likely dynamic in the data, our cross-sectional data prevents us from observing such dynamics. Our focus is therefore on how subjective literacy impacts housing choices. As discussed in Section 4.3.1, our estimation strategy allows us to answer this question while largely abstracting from the dynamics of financial literacy.

[7]For robustness, we also consider additional variations of the regression Equation 1, including incorporating the continuous version of our literacy measures and interacting this measure with other control variables to allow for differential relationships with respect to housing market outcomes. Results are quantitatively similar. These results are also largely robust to the way we divide households into literacy categories.

Income

Households face both idiosyncratic and aggregate income shocks. In each period until retirement at age $a = Ret$, households are endowed with labor income Y_t that follows an exogenous stochastic process. Following Cocco, Gomes and Maenhout (2005), the income process before retirement is given by:

$$logY_t = f(a_t) + log\overline{Y_t} + log\hat{Y}_t^i + u_t, \tag{2}$$

where $f(a_t)$ is a deterministic life cycle profile and u_t is an idiosyncratic temporary shock distributed as $N(0, \sigma_u^2)$. $\overline{Y_t}$ and $\hat{Y}_t^i$ are the aggregate and idiosyncratic components of income, both following a random walk in logs:

$$log\overline{Y_t} = log\overline{Y_{t-1}} + \overline{\epsilon_t}$$
$$log\hat{Y}_t^i = log\hat{Y}_{t-1}^i + \hat{\epsilon}_t^i,$$

where $\overline{\epsilon_t}$ is distributed $N(0, \sigma_{\overline{\epsilon}}^2)$ and $\hat{\epsilon}_t^i$ is distributed $N(0, \sigma_{\hat{\epsilon}}^2)$. The shocks $\overline{\epsilon_t}, \hat{\epsilon}_t^i, u_t$ are uncorrelated. These assumptions allow us to denote the permanent shock to household income as:

$$\epsilon_t^{\hat{Y}} = \overline{\epsilon_t} + \hat{\epsilon}_t^i \sim N(0, \sigma_{\hat{Y}}^2).$$

Following retirement, households receive a constant fraction θ_{Ret} of their income in the period prior to retirement.

Preferences and Choices

Each period, households choose the amount of housing services S_t and numeraire consumption C_t. Lifetime utility is given by:

$$E_0 \left\{ \sum_{t=0}^{A} \beta^t \left[\left(\prod_{j=0}^{t} \lambda_{a_j} \right) \lambda_{a_{t+1}} u(C_t, S_t) + \left(\prod_{j=0}^{t} \lambda_{a_j} \right) (1 - \lambda_{a_{t+1}}) D_t \right] \right\},$$

where D_t is the bequest utility in case of death and $u(C_t, S_t)$ is the per-period utility.

Households can consume housing services in two ways: by renting or by owning a house. Denote by $\tau_t \in \{0, 1\}$ the tenure choice at time t, with $\tau_t = 1$ indicating ownership. A house of quality H_t provides housing services according to the linear technology:[8]

[8]Many models of portfolio choice with housing incorporate an age-dependent preference for tenure which is driven by exogenous forces such as uncertainty regarding changes in workplace and household size. Following Landvoigt (2017), we also solve a specification of the model where $S_t = \phi(\tau_t, a_t) H_t$ and $\phi(\tau_t, a_t) = 1 + (1 - \tau_t) e^{-\kappa a_t}$. κ then regulates the age-dependent preference to own. Since our baseline model fits the housing market data patterns, we proceed without incorporating an age-dependent preference.

$$S_t = H_t.$$

The functional form of the per-period utility function is the standard Cobb-Douglas:

$$u(C_t, S_t) = \frac{\left[C_t^{\rho} S_t^{1-\rho}\right]^{\gamma}}{1-\gamma},$$

where γ is the relative risk aversion parameter and ρ measures the intra-temporal substitution between housing and other consumption goods. The bequest utility D_t is a function of the household's total wealth in period t, W_t, as well as house prices, and is given by:

$$D_t(W_t, P_t) = \frac{\overline{D}(W_t^i / P_t^{\rho})^{1-\gamma}}{1-\gamma},$$

where $\overline{D}$ mediates the importance of bequest motives relative to other consumption. The functional form of the bequest function, namely the normalization by house prices, is chosen to ensure value function homogeneity.

Houses and Prices

Households can rent each *quality unit* of housing for a price P_t^r. The per-period cost of renting a house of quality H_t is therefore $P_t^r H_t$. For homeowners, houses serve not only as a consumption good but also as an asset. Each *quality unit* of the housing asset sells for a price of P_t. The house price of a house of quality H_t is therefore $P_t H_t$. House prices are subject to aggregate risk. Specifically, the price per quality unit of housing follows a random walk in logs:

$$\log(P_t) = \log(P_{t-1}) + \epsilon_t^P,$$

where $\epsilon_t^P \sim N(d_P, \sigma_P^2)$ and d_P is the deterministic drift in house price growth. We assume that the vector of shocks to income and house prices $(\epsilon_t^{\hat{Y}}, \epsilon_t^P)$ is independent across time with a variance matrix of:

$$Var(\epsilon_t^{\hat{Y}}, \epsilon_t^P) = \begin{bmatrix} \sigma_{\hat{Y}}^2 & \sigma_{\hat{Y}P} \\ \sigma_{\hat{Y}P} & \sigma_P^2 \end{bmatrix}.$$

Aggregate shocks to the price per quality unit of housing might hence be contemporaneously correlated with permanent shocks to income.

House prices are also subject to idiosyncratic risk. Specifically, the quality of an owner-occupied home, H_t, is itself stochastic and evolves according to the idiosyncratic process:

$$H_{t+1} = Q_i(H_t) = (1 + g_{i,t+1})H_t,$$

where $g_{i,t} \sim N(\mu(f_i), \sigma^2(f_i))$ is i.i.d across time. Thus, the evolution of the house price, $P_t H_t$, depends on (1) the aggregate shock to the price per quality unit of housing and

(2) on the idiosyncratic shock to the quality of housing. The latter can depend on the subjective financial literacy of the household that owns it. It is useful to note that the return on a house that is owned by household i is given by:

$$\frac{P_{t+1}H_{t+1} - P_t H_t}{P_t H_t} = \exp\left(\epsilon^P_{t+1}\right)\left(1 + g_{i,t+1}\right) - 1.$$

Our model echoes the idea that households with higher financial literacy may have access to better investment opportunities in the housing markets, due to, e.g., sophisticated search skills. We model this form of heterogeneity by allowing both the mean and volatility of the distribution of idiosyncratic shocks to house prices to differ by literacy. Note that in the baseline model, expectations on future house prices are aligned with the true distribution of returns. That is, to the extent that households with different levels of subjective financial literacy hold different expectations, this reflects true fundamental differences in investment opportunities. In an alternative specification of the model, we consider a case where heterogeneous expectations on future house prices reflect differences in beliefs across households with different self-assessed literacy, but where realizations of returns are drawn from a common distribution (Section 5.3). That is, beliefs might be distorted. We find that the benchmark model is better at explaining the housing choices in the data, suggesting that subjective financial literacy reflects, at least in part, true financial savviness.

Collateral Constraints and Default

Households are allowed to save in a risk-free asset which generates R units of return at $t+1$ for each unit of the numeraire saved in t. When borrowing, households pay a financial-literacy dependent interest rate spread of $\varrho(f_i) > 0$, appealing to the possibility that financially literate households might search and negotiate for cheaper credit. Borrowing is also subject to a collateral constraint. Only homeowners can borrow, and they can borrow up to a ratio of $(1 - \delta(f_i))$ of the value of their house. The collateral constraint can therefore vary across different levels of financial literacy, alluding to the possibility that financially literate households might have access to larger credit. Denote by $B_t \geq 0$ savings and by $B_t < 0$ borrowing. The collateral constraint is given by:

$$B_t \geq \begin{cases} 0 & \tau_t = 0 \\ -\left[1 - \delta(f_i)\right] P_t H_t & \tau_t = 1 \end{cases}. \tag{3}$$

Budget Constraints

When specifying the budget constraint, we distinguish between two cases: households that have rented in the previous period and households that were owners in the previous period. For simplicity, we assume the rental price per quality unit of housing is pegged to the selling price per quality unit, that is $P_t^r = \alpha P_t$.

Case 1: Previous Renters

The time t budget constraint for a household that was renting in period $t-1$ is given by:

$$C_t + B_t + P_t H_t \Big\{ (1-\tau_t)\alpha + \tau_t \ (1+\psi) \Big\} = RB_{t-1} + Y_t, \tag{4}$$

where ψ accounts for the proportional maintenance cost that an owner must incur every period to offset depreciation. A previous renter enters the period with total wealth (or "cash-on-hand") W_t, which is the sum of accrued savings and contemporary income. It chooses how much to consume, how much to save in bonds, whether or not to purchase a house (in which case it can also borrow), and how much housing to consume.

Case 2: Previous Owners

Previous owners choose whether or not to sell their house. If they sell, they choose whether to rent or own and how much housing to consume. We denote the decision of whether to sell or not by $\xi_t = \{0, 1\}$, where $\xi_t = 1$ indicates selling. A previous owner who chooses to sell faces the following budget constraint:

$$\begin{aligned} C_t + B_t + P_t H_t \Big\{ (1-\tau_t)\alpha + \tau_t \ (1+\psi) \Big\} = \\ \Big[R + 1_{\{B_{t-1}<0\}} \varrho(f_i) \Big] B_{t-1} + Y_t + (1-\nu) P_t \left(1 + g_{i,t}\right) H_{t-1}, \end{aligned} \tag{5}$$

where ν accounts for the proportional transaction cost that a seller incurs. A previous owner who chooses not to sell faces the following budget constraint:

$$C_t + B_t + \psi P_t \left(1 + g_{i,t}\right) H_{t-1} = \Big[R + 1_{\{B_{t-1}<0\}} \varrho(f_i) \Big] B_{t-1} + Y_t. \tag{6}$$

The housing services for this household is $S_t = \left(1 + g_{i,t}\right) H_{t-1}$. Finally, previous owners might be hit by an exogenous moving shock, in which case they are forced to sell their house. Moving shocks are i.i.d and drawn from a distribution that can depend on age. Moving shocks capture life-cycle shocks that induce selling and which are not captured by the model. We denote the moving shock by M_t, where $M_t = 1$ indicates that the household is forced to moved.

Bellman Equations

The recursive nature of the problem allows us to state it in terms Bellman equations. Denote by $X_t = \{a_t, f_i, W_t, P_t, \tau_{t-1}, (1+g_{i,t})\,H_{t-1}, \hat{Y}_t, M_t\}$ the vector of household state variables where $W_t = \left[R + 1_{\{B_{t-1<0}\}}\varrho(f_i)\right] B_{t-1} + Y_t$, and $P_t\,(1+g_{i,t})\,H_{t-1}$ is the realized house price that owners can sell their house for. In addition, denote by $Z_t = \{\tau_t, H_t, C_t, B_t, \xi_t\}$ the household vector of choices. The following problem specifies the household value function for households of age $a < Ret - 1$:[9]

$$
\begin{aligned}
&V(a_t, f_i, W_t, P_t, \tau_{t-1}, (1+g_{i,t})\,H_{t-1}, \hat{Y}_t, M_t) = \\
&\lambda_{a_t} \max_{Z_t} \left\{ u(C_t, S_t) + \beta \mathbb{E}_t^i [V(a_t+1, f_i, W_{t+1}, P_{t+1}, \tau_t, (1+g_{i,t+1})\,H_t, \hat{Y}_{t+1}, M_{t+1})] \right\} + \\
&(1-\lambda_{a_t}) D(W_t, P_t),
\end{aligned} \tag{7}
$$

where $\hat{Y}_t = \overline{Y_t}\hat{Y}_t^i$ is the permanent income component. The problem is subject to the collateral constraint (Equation 3) and budget constraint (Equations 4-6).

The household problem can be solved by employing standard dynamic programming methods. In order to reduce the state space dimensionality and efficiently compute the policy functions, Appendix B presents a normalized and equivalent problem. The solution follows Landvoigt (2017) and relies on the homothetic nature of the problem.

3.2 Discussion

Our goal is use the model to quantify the role of expectations and mortgage terms in explaining the empirical relationship between subjective financial literacy and housing choices (Section 2). Note that our model is a partial equilibrium one - we do not model the supply side of the economy or solve for equilibrium prices in the housing market. Rather, the model solution yields households' optimal tenure and mortgage choices as a function of their state. We estimate the model parameters, namely 1) the expected idiosyncratic shock to homeowners' house prices, $\mu(f_i)$; 2) the expected volatility of this shock, $\sigma(f_i)$; 3) the mortgage interest rate spread, $\varrho(f_i)$; and 4) the minimum collateral requirement,

[9]The Bellman equations for $a \geq Ret - 1$ are given in Appendix B.

$\delta(f_i)$, so that optimal housing choices in the model match the housing choices that we observe in the data. In what follows, we discuss which data moments are most important for the identification of each of these parameters.[10]

Homeownership and Credit Conditions

Tenure decisions in the model are primarily driven by mortgage market parameters. The collateral constraint is particularly important for explaining tenure choices of young households. Intuitively, households with little wealth, and who tend to be younger, need to borrow in order to buy a house. The collateral constraint governs the extent to which they are able to do so. A stringent collateral requirement screens young households out of the owner-occupied market. Moreover, young households are more prone to borrowing since the deterministic life-cycle component of their income is upward sloping. However, the collateral constraint limits their ability to borrow. Relaxing the constraint therefore particularly impacts the young. To sum, the collateral requirement, $\delta(f_i)$ is mostly identified from differences in ownership rates across young households with varying degrees of subjective financial literacy.

Mortgage spreads are more important for the tenure choices of middle-aged and older households. As in standard quantitative life-cycle models, the deterministic component of household income in the model is hump-shaped. This means that middle-aged and old households expect their future income to decrease and are therefore prone to save. However, many of these households have yet to pay their off the mortgages that they took on when they were younger. The extent to which they are willing to continue paying off their debt instead of saving (by selling their house and moving into a rental or downsizing to a lower quality owner-occupied home), depends on how expensive it is to continue owning. This in turn depends on how large the mortgage spread is. The mortgage spread, $\varrho(f_i)$ is therefore mostly identified from differences in ownership rates across older households with varying degrees of subjective financial literacy.

Leverage and Expectations

While credit conditions are mostly identified by tenure decisions, expectations on future house prices are mostly identified by leverage choices. Houses serve not only as a consumption good but also as an asset that households can save in. Conditional on choosing to buy a house of a certain value, households are more likely to lever more when they

[10]Parameters are jointly estimated to match data moments using Simulated Method of Moments (SMM). Nevertheless, it is useful to relate each parameter to the data target it affects most quantitatively.

expect the idiosyncratic shock to house price growth to be higher and less volatile. The relative importance of the volatility of the shock vis-à-vis its expected mean depends on households' age. Namely, older households are relatively more sensitive to the volatility parameter. This is because of their lower net present value of the non-risky component of income, which makes their optimal portfolio choice more sensitive to increases in risk. This is in contrast to younger households', for whom the higher present value of the non-risky component serves as a hedge. For these households, it is optimal to take on risk even when the volatility is higher. To sum, the expected idiosyncratic shock to house price for homeowners, $\mu(f_i)$, and the expected volatility of this shock, $\sigma(f_i)$, are mostly identified from loan-to-value ratios of homeowners with varying degrees of subjective financial literacy.

4 Quantification

We quantify the model to the U.S. housing markets. It is helpful to group parameters into two categories: those that are calibrated exogenously, and those that are estimated internally to match the empirical relationship between subjective financial literacy, housing tenure, and mortgage choices.

4.1 Data

We quantify the model using the 2016 cross-section of the SCF. The data include information on balance sheets, income, and demographic characteristics of a representative sample of U.S. households. As discussed in Section 2, the measurement of subjective financial knowledge was first introduced to the 2016 questionnaire, limiting us to the use of this particular wave. We use the summary extract public data of the SCF and focus on families for which the head of household is aged between 25 and 80, the life-span considered in our model. Total wealth is defined by the SCF as the balance between total assets (financial and non-financial) and total debt, coded as "networth". We omit households with total net-worth larger than 7 million dollars. Our model is not suitable for describing the life-cycle wealth dynamics of rich households who rely much more on stock market capital gains and non-traditional retirement income sources. Applying the SCF sampling weights, these households consist of about 11% of effective observations[11]. We define total labor income as the sum of wage income, income from retirement and social security funds, income from self managed businesses and transfers from other sources. We use

[11] Abstracting from the stock market in the model thus seems reasonable for the lower 90% of households.

the variable "houses" as the value of the house (for owners) which is defined by the SCF as the value of the primary residence. Mortgage debt for homeowners is coded by the SCF as "mrthel" and includes all forms of debt which are collateralized against the value of the house. Subjective financial knowledge is measured by "knowl" and is categorized into three groups, as discussed in section 2.

4.2 Calibration

The model is calibrated at annual frequency. Households enter the model at age 25 and live until the age of 80. Table 3 reports the model parameters that we calibrate exogenously. The preference parameters are taken from the literature. Risk aversion γ is set to 3 and the Cobb-Douglas weight on housing services, ρ, is set to 13% based on Piazzesi, Schneider and Tuzel (2007). The strength of the bequest motive is estimated internally and discussed below.

The income process is calibrated based on Cocco, Gomes and Maenhout (2005). The deterministic part $f(a_t)$ follows a three-degree polynomial in age. We use the coefficients characterizing the life-cycle profile of high-school graduates estimated by Cocco, Gomes and Maenhout (2005) from PSID data, and adapt them to fit our income specification. The life-cycle component has the usual hump shape. The annual standard deviation of the permanent shock is set to 10.6%. The correlation between innovations to house price and permanent income, $\sigma_{\hat{Y}P}$, is set to zero based on Flavin and Yamashita (2002). θ_{Ret} is set to be 0.7 in accordance with Cocco, Gomes and Maenhout (2005).

Moving to prices, the risk-free interest rate R is calculated as the average real yield of a 1-year treasury bond between 2010-2019. We set the drift in house price growth to be equal to that implied by the estimated income process, which is 0.5%. This ensures that the ratio between house prices growth and income growth is stationary. The volatility of house price growth, σ_P^2, is estimated internally and discussed below. To compute the rent-to-price ratio α we use the FHFA aggregate price index and deflate it by the CPI of house rental prices. The long run value of this series is consistent with Davis et al. (2008) and Sommer et al. (2013). The maintenance cost accrued by homeowners in order to offset depreciation is set as a 1% share of the house value, and is line with other values in the housing literature.

We estimate age-dependent moving probabilities using the 2010 Census data. To identify moving for reasons that are exogenous to our model (e.g. marriage, divorce) we use the 2015 American Housing Survey which asks respondents for moving circumstances. The life-cycle mobility shock is estimated to be downward sloping with age. Finally, survival rates λ_a are calculated from The National Center of Health Statistics mortality rates.

Table 3: Exogenous Parameters

Parameter	Notation	Value
Relative risk aversion	γ	3
Housing services weight in utility	$1-\rho$	0.13
Relative income at retirement	θ_{Ret}	0.7
Permanent shock volatility	$\sigma^2_{\hat{\gamma}}$	0.0106
Transitory shock volatility	σ^2_u	0.0738
Risk-free rate	R	1.01
Drift in house price growth	d_P	0.005
Maintenance cost	ψ	0.01
Rent to price ratio	α	0.05
Transaction cost	ν	0.08

4.3 Estimation Procedure

4.3.1 Simulated Method of Moments Approach

We estimate the remaining model parameters by applying a Simulated Method of Moments (SMM) to the cross-sectional 2016 SCF data. Our estimation strategy is discussed in detail in Appendix C. In what follows, we provide a brief summary.

Denote the set of parameters to be estimated, which we specify below, by the vector η. We begin by simulating a large number of I households from the SCF data in 2016. For each sampled household, we observe the vector of its (normalized) state variables. Given these state variables, given the exogenously calibrated parameters, and given a guess for η, we obtain each household's optimal policies by solving the household problem. For each household, we then draw the permanent and transitory shocks to income, the aggregate shock to the price per quality unit of housing, and the idiosyncratic shock to house price growth. Together with the household policies, this maps the sample of simulated households in 2016 to a sample of simulated households in 2017. We then estimate η by minimizing (in an SMM fashion) the distance between the sample of simulated households in 2016 and the sample of simulated households in 2017.

The estimation relies on two assumptions. First, since our data is not of panel structure, the observed households in 2016 are not followed into 2017. That is, we do not observe the 2017 sample in the SCF data. Comparing the simulated samples in 2017 and the simulated sample in 2016 thus assumes that the 2016 sample represents an invariant distribution of households (up to the secular growth of prices and income). The second assumption is that subjective financial literacy is innate. That is, each household's subjective financial literacy in 2017 is the same as it is in 2016 . While literacy is likely dynamic in the data, we are unable to observe such dynamics since our data is cross-sectional. However, even if literacy is indeed dynamic, this would pose a concern to our estimation only to the extent that it evolves between two consecutive years. That is, the estimation only requires assuming that literacy is innate within a one-year period. Appendix C specifies the estimation procedure in more detail and reports standard errors. In what follows we discuss the data moments we target and the parameters we estimate.

4.3.2 Parameters and Moments

The parameters that we estimate internally include all the parameters that depend on subjective financial literacy: 1) the expected idiosyncratic shock to house prices for homeowners, $\mu(f_i)$; 2) the expected volatility of this shock, $\sigma(f_i)$; 3) the mortgage interest rate spread, $\varrho(f_i)$; and 4) the minimum down-payment requirement, $\delta(f_i)$. The data moments we target in the SMM estimation are homeownership and loan-to-value moments. Specifically, using the SCF data, we compute the homeownership rate and loan-to-value for young households (those between the ages of 25 and 40), middle aged households (between the ages 41 and 60), and old households (those older than 60). Each of these moments is further broken down by the three types of financial literacy. Overall, this gives 12 parameters and 18 moments.

In addition to the financial literacy dependent parameters, we also estimate the discount factor β to match aggregate wealth in the data, the strength of bequest motives $\overline{D}$ to match the average wealth at age 80, and the growth volatility of the price per-quality-unit of housing, σ_P^2, so that the growth volatility of house prices in the model is 15%. This number reflects both idiosyncratic risk, which Landvoigt et al. (2015) and Case and Shiller (1990) estimate to be between 9% and 15%, and aggregate housing risk which Flavin and Yamashita (2002) estimate to be between 5% and 9%.[12]

5 Results

Table 4 reports the estimation results. The results suggest that households that self-assess themselves as more literate face laxer constraints in the credit markets - they pay a lower

[12]In an additional exercise, we also allow for heterogeneity in the discount factor across self-assessed financial literacy groups. Data moments in this case are augmented with wealth by literacy groups. Such heterogeneity doesn't seem to play an important role, as estimates are basically equal across levels of self assessed literacy.

spread when borrowing against the value of their house, and are subject to a more lenient collateral constraint. Relative to households with low subjective literacy, those who view themselves as highly literate face a 3 percentage points lower mortgage spread, and can borrow about 6 precent more against the value of their house.

Households with higher subjective financial literacy also have more optimistic expectations on future house price growth. Relative to households with low subjective financial literacy, households with high subjective financial literacy expect the idiosyncratic shock to house price growth to be drawn from a distribution with a higher mean and lower volatility. While both the expected return and the volatility of the idiosyncratic shock to house price growth is highest for households with intermediate levels of subjective financial literacy, the coefficient of variation (CV), which measures how risky the investment is, is decreasing with subjective financial literacy.

Table 4: Internally Estimated Parameters

Parameter	**Low Literacy**	**Intermediate Literacy**	**High Literacy**
Expected return $\hat{\mu}(f)$	0.04 (0.0004)	0.07 (0.0005)	0.05 (0.0001)
Volatility $\hat{\sigma}(f)$	0.037 (0.0005)	0.058 (0.005)	0.031 (0.002)
Coefficient of Variation $\frac{\hat{\sigma}(f)}{\hat{\mu}(f)}$	0.935	0.833	0.588
Mortgage spread $\hat{\varrho}(f)$	0.04 (0.0002)	0.028 (0.002)	0.009 (0.0006)
Min. down-payment $\hat{\delta}(f)$	0.2 (0.005)	0.154 (0.0005)	0.143 (0.01)

Notes: Parameters are estimated by SMM , as described in Section 4.3.1. Standard errors, in parenthesis, are discussed in Appendix C.

5.1 Model Fit

Figure 2 shows the fit of our model to the data. The model is able to closely match the stylized facts. As in the data, model-implied homeownership rates and loan-to-value ratios are increasing with subjective financial literacy. Both in the data and in the model, homeownership rates exhibit a steep slope in financial literacy for all age groups. As discussed in Section 3.2, the differences in homeownership rates across young households with varying subjective financial literacy mostly identify the heterogeneity in collateral requirements in the model. Differences in ownership rates across middle-income and older households mostly identify the heterogeneity in mortgage spreads in the model. Both in the model and in the data, differences across levels of financial literacy are less stark when considering loan-to-value ratios, and shows up only for middle-aged and old households. These leverage differences mostly identify heterogeneity in the expected mean of the idiosyncratic shock to house price growth and the expected volatility of this shock.

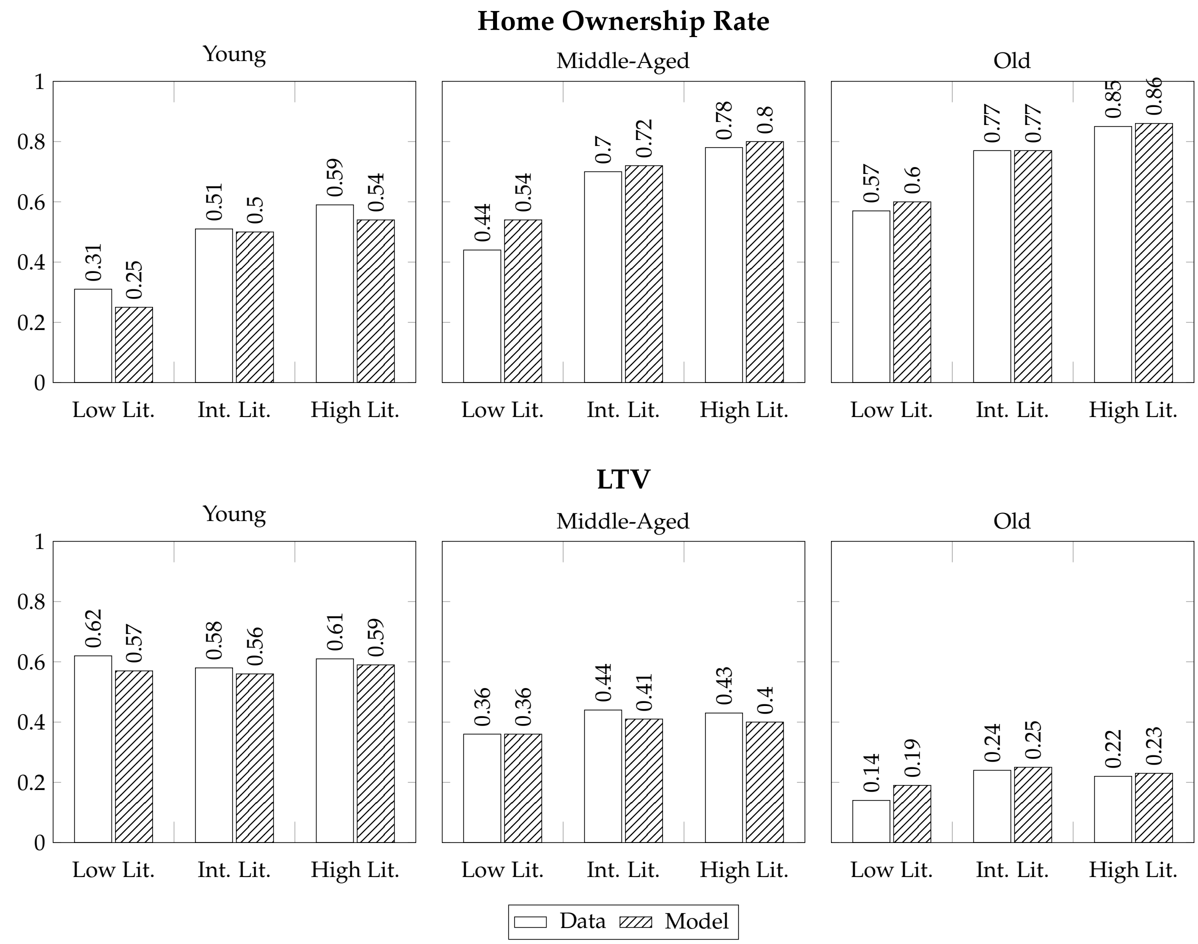

Figure 2: Full Model: Target and Model Generated Moments

Note: The figure compares model generated moments to SCF data moments. The Loan-to-Value ratio is averaged across all homeowners and is computed in the SCF as the ratio of all debt which is collateralized against the house, divided by the value of the house . Young households are those in which the household head is 40 years old or younger, middle-aged are those between the ages of 41 and 60, and the old are those older than 61. Low literacy households are those self-assessing their knowledge to be between 1 and 4 on the $1-10$ scale, intermediate literacy households are those self-assessing their knowledge to be between 5 and 7 and high literacy households are those self-assessing their knowledge to be between 8 and 10.

By fitting the data, our model suggests that heterogeneity in mortgage terms and in expectations on future house prices can account for the empirical relationship between subjective financial literacy and housing choices. While there might be additional potential channels that can rationalize our empirical findings, our focus is on those that are arguably the most intuitive - mortgage terms and housing market expectations. We further validate our model by showing that it also matches non-targeted moments that are important for the relationship between subjective financial literacy and housing choices. Namely, while the model targets the relationship between financial literacy and housing choices *unconditional* on income and wealth, it also closely matches the *conditiona*l correlation. This is illustrated in Table 7, which regresses tenure and leverage choices in the model and in the data, controlling for age, wealth and income. We discuss these results in more detail below.

5.2 Mechanisms

We have thus far shown that households that self-assess themselves as more financially literate face laxer mortgage terms and expect better risk-return trade-offs in housing markets. But how important are each of these two mechanisms in generating the documented stylized facts? To answer this question, we consider two variants of our model. In the first, we shut off heterogeneity in expectations and continue to allow heterogeneity in mortgage markets In the second, we consider the analog case where only heterogeneity in expectations is allowed. We then ask how the fit of these models to the data compares to the fit of the full model that allows heterogeneity along both dimensions.

We begin by evaluating a model in which households with different subjective financial literacy have access to different mortgage terms but have the same expectations on future house prices. Namely, we simulate a model where we maintain the estimates of mortgage spreads $\varrho(f_i)$ and minimum collateral requirement $\delta(f_i)$ from Table 4, but set the expected returns on the idiosyncratic shock to house prices, μ, and the expected volatility of the shock, σ, to be equal across literacy types. Specifically, we use the average of $\mu(f_i)$ and $\sigma(f_i)$ from Table 4 (weighted by the relevant population shares). The results of this exercise are illustrated in Figure 4 in the appendix.

The main takeaway is that when heterogeneity in expectations are shut off, the fit of the model with respect to the data becomes worse for the middle-aged and old, but actually slightly improves for the young. This can be seen in both housing market outcomes, and across the three literacy types. This suggests that expectations matter for explaining the link between subjective financial literacy and housing choices among older households, but less so for explaining the variation among young households. This is intuitive. Ownership and leverage decisions of older households, who are less likely to be borrowers, are mostly driven by the risk-return they expect in housing markets. Thus, when heterogeneity in these expectations is shut off, the model's ability to match the differences in housing choices across older households with different subjective literacy is dampened. In contrast, the model ability to match differences across younger households is unharmed, since younger households' housing choices mostly depend on borrowing conditions.

Next, we evaluate a model in which households with different subjective financial literacy have different expectations on future house prices but face the same mortgage terms. Namely, we simulate a model where we maintain the estimates $\mu(f_i)$ and $\sigma(f_i)$ from Table 4, but set ϱ and δ to their weighted average values. The results of this exercise are given in Figure 4 in the appendix. When heterogeneity in mortgage markets is shut off, the fit of the model with respect to the data deteriorates slightly more for younger and middle-aged households relative to older households. For example, for the middle-aged, this model does worse in terms of matching both homeownership and loan-to-value for the low-literacy households as well as the loan-to-value ratio for the high literacy group. At the same time, for the old households there doesn't seem to be much of a difference between the two models. The results suggest that mortgage terms matter more for explaining the link between subjective financial literacy and housing choices among younger households. Intuitively, since ownership and leverage choices for younger households depend more on access to credit, when heterogeneity in credit conditions is dismissed, the model's ability to match differences in housing choices across young households is dampened. For older households, such heterogeneity matters less since they are less prone to borrowing. Finally, note that when heterogeneity in credit conditions is shut off, the fit of the model with respect to the data deteriorates relatively less compared to when heterogeneity in expectations is dismissed. This suggests that, overall, expectations might be more important in explaining the observed empirical patterns.

5.3 Subjective or Objective Expectations?

The results thus far suggest that heterogeneity in expectations might play an important role in explaining why households that self-assess themselves as more financially literate are also more likely to own and take on more leverage. In our baseline model, we have assumed that subjective financial literacy proxies true financial savviness. Namely, we have assumed that expectations on idiosyncratic shocks to house prices are aligned with the true distributions from which these shocks are drawn. To the extent that households with different subjective financial literacy hold different expectations, in the baseline model this reflects true - objective - differences in investment opportunities (for example, due to more sophisticated search skills).

An alternative view is that subjective literacy proxies distorted beliefs. For example, households that self-assess themselves to be more financially literate might be over-optimistic (or over-pessimistic). In an extension of our main analysis, we test the role that distorted beliefs might play in explaining the empirical facts. To do so, we consider a specification of the model where we allow for heterogeneous expectations on idiosyncratic shocks to house prices, but in which the actual distribution from which these shocks are drawn is independent of financial literacy. That is, households with different levels of financial literacy solve different maximization problems, based on their individual beliefs $\{\mu(f_i), \sigma(f_i)\}$, but the realizations of $g_{i,t}$ are drawn from the same distribution. The common distribution from which we draw $g_{i,t}$ is normal with the mean and variance set to their weighted average values from Table 4.

The fit of this model to the data is illustrated in Figure 6 in the appendix. When heterogeneity in the distribution of realized returns is shut down, the fit of the model with respect to the data deteriorates relative to the baseline model that admits such heterogeneity. This is seen mostly in terms of loan-to-value ratios and to some extent also in terms of homeownership rates. The main takeaway from this analysis is that self-assessed financial literacy proxies, at least to some extent, true - objective - financial literacy. Heterogeneity in the fundamental distribution of idiosyncratic shocks to house prices improves the model's ability to match the data.

6 Financial Literacy and Housing Policies

We have thus far shown that 1) more financially literate households are more likely to own and take on larger mortgages, and 2) this can be accounted for by heterogeneity in mortgage terms and expectations on future house prices. In this section, we ask how important is heterogeneity in financial literacy for the evaluation of housing policies. To answer this question, we compare our model to a benchmark portfolio choice model with housing where heterogeneity in financial literacy is abstracted from. We then compute the impact of counterfactual housing policies in both models. This exercise allows us to quantify the potential bias in policy evaluation that arises from abstracting from heterogeneity in financial literacy.

6.1 Benchmark Model without Heterogeneity

We begin by estimating a benchmark model in which heterogeneity in financial literacy is muted. That is, we restrict all parameters to be independent of financial literacy. The data moments we target are the same as in the estimation of the full heterogeneous agent model (Section 4.3.1), with the caveat that we now compute the moments unconditional on financial literacy. The estimation results for this benchmark model are reported in Table 5.

Table 5: Estimated Parameters: Benchmark Model

Parameter	Estimated Value
Expected return $\hat{\mu}$	0.07 (0.0003)
Volatility $\hat{\sigma}$	0.064 (0.0005)
Coefficient of Variation $\frac{\hat{\sigma}(f)}{\hat{\mu}(f)}$	0.855
Mortgage spread $\hat{\varrho}$	0.028 (0.0002)
Min. down-payment $\hat{\delta}$	0.17 (0.0001)

Notes: Parameters are estimated by SMM , as described in Section 4.3.1. Standard errors, in parenthesis, are discussed in Appendix C.

Model Fit. The benchmark model closely matches the life cycle dynamics of homeownership rates and loan-to-value ratios. Figure 3 shows this by plotting the model generated moments against the data moments. Given that this benchmark model successfully accounts for the life cycle dynamics we see in the data, why should we consider augmenting it with heterogeneity in financial literacy? In what follows, we show that ignoring this dimension of heterogeneity results in substantial biases when evaluating housing market policies.

6.2 Housing Policies With and Without Financial Literacy

In this section, we consider the effect of a wealth shock on housing choices. We compare the effect of the shock in the benchmark model to its effect in the model that allows for heterogeneity in financial literacy. The wealth shock proxies policies that are designed to encourage homeownership, for example income transfers, tax deductions or capital gains exemptions. Our main finding is that the impact of a wealth shock on homeownership is downsized by approximately 40% in a model with heterogeneity in financial literacy. For example, for young households (age 25-40), we find that a 10% increase in wealth leads to a 10% increase in homeownership in the benchmark model, but only to a 6.4% increase in homeownership in the heterogeneous agent model. For young and poor households (age 25-40 and in the bottom quantile of the wealth distribution), a 10% increase in wealth leads to a 20% increase in homeownership in the benchmark model, but only to a 11% increase in homeownership in the heterogeneous agent model. The corresponding housing demand elasticities are reported in Table 6. Overall, the main takeaway is that the eval-

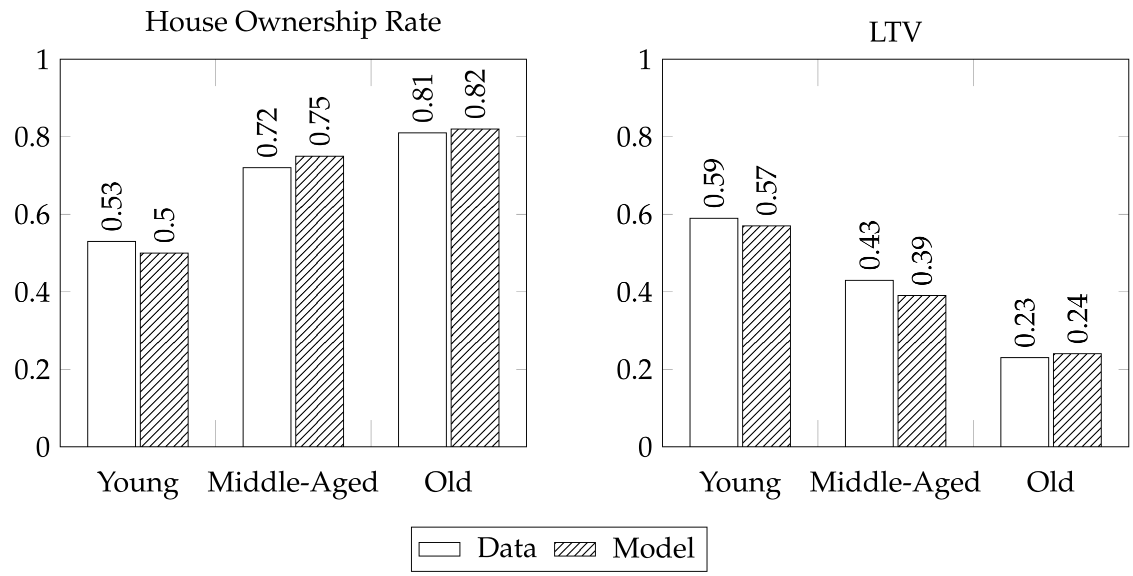

Figure 3: Benchmark Model: Target and Model Generated Moments

Note: The figure compares moments generated by the benchmark model (without financial literacy heterogeneity) to SCF data moments. The Loan-to-Value ratio is averaged across all home-owners and is computed in the SCF as the ratio of all debt which is collateralized against the house, divided by the value of the house. Young households are those in which the household head is 40 years old or younger, middle-aged are those between the ages of 41 and 60, and the old are those older than 61.

uation of housing market policies crucially depends on whether or not heterogeneity in financial literacy is taken into account.

Table 6: Housing Demand Elasticity - With and Without Financial Literacy

Population	Benchmark Model	Heterogeneous Model
Young	1	0.64
Young and Poor	2	1.1

Notes: Housing demand elasticity is computed as the percent increase in home-ownership in response to a one percent increase in wealth. The first row shows this elasticity for young households (aged 25-40) whereas the second row focuses on young and poor household (from the bottom quantile of the wealth distribution).

Next, we discuss the underlying reason for the discrepancy between the benchmark model and the heterogeneous agent model. The key finding is that the benchmark model over-estimates the correlations between housing outcomes and wealth, income, and age relative to the data. This in turn leads to biased evaluations of housing policies. The heterogeneous agent model, in contrast, is able to substantially reduce the bias in the correlation between ownership and wealth, income, and age relative to the data. As a result, it produces more reliable policy evaluations.

To see this, we estimate the following regression specification in the data, in the benchmark model, and in the heterogeneous agent model:

$$Y_i = \beta_{low} FK_{low,i} + \beta_{high} FK_{high,i} + \Gamma X_i + \epsilon_i. \tag{8}$$

Controls X_i include household age, age-squared, wealth-to-income and wealth-to-income quartiles. $FK_{low,i}$ ($FK_{high,i}$) is an indicator equal to one if the household belongs to the low (high) financial literacy groups. The results are reported in Table 7.[13]

Compared to the data (column 1), the benchmark model (column 2) generates an excessive co-movement of homeownership with wealth-to-income and with age. A one percent increase in the wealth-to-income ratio is associated with a 0.928 increase in the homeownership log-odds ratio in the data, but a 5.87 increase in the benchmark model. Similarly, a one-year increase in age is associated with a 0.031 increase in the homeownership log-odds ratio in the data, but a 0.4 increase in the benchmark model. A one percent increase in wealth-to-income is associated with a 18.4% reduction in loan-to-value in the benchmark model (column 5), compared to only a 9% decline in the actual data (column 4).[14]

The heterogeneous agent model significantly reduces these biases. As Table 7 shows, across the board, the model implied regression coefficients converge towards the data when heterogeneity in financial literacy is introduced. For example, a one percent increase in the wealth-to-income ratio is associated with only a 2.55 increase in the homeownership log-odds ratio in the heterogeneous agent model (column 3), much lower than the 5.87 increase in the benchmark model (column 2), and much closer to the 0.928 estimate in the data (column 1). It now becomes clear that, by overstating the correlation between wealth and housing choices, the benchmark model delivers inflated evaluations of the elasticity of housing demand relative to wealth. By more accurately capturing these correlations, the heterogeneous agent model therefore delivers lower, and less biased, policy evaluations.

We should point out that adding a new source of heterogeneity in any dimension will mechanically reduce the excessive correlations between wealth and housing outcomes

[13]Comparing the model-generated estimates to those from the SCF data requires estimates and standard errors be computed in a similar fashion. As discussed in Section 2, in order to accommodate for the complex sampling design of the SCF, estimates and standard errors are computed by applying a bootstrapping routine. We therefore follow this routine for computing the model-implied estimates. We draw 1,000 bootstrap samples from the the SCF distribution of the model state variables. We then apply the policy functions on each sample to simulate model-implied regression estimates.

[14]Financial literacy in the benchmark model is correlated with homeownership and loan-to-values, despite the fact that parameters do not differ across literacy levels. The reason is that financial literacy in the data (and therefore in the model simulation) is correlated with house values and persistent income, which are state variables in the model.

Table 7: Data and Model Regressions

	Home Ownership			LTV		
	Data	Bench.	Full	Data	Bench.	Full
Low Fin. Lit.	-0.644^{***} (0.119)	-0.206^{***} (0.071)	-0.711^{***} (0.035)	-0.078^{**} (0.027)	0.00 (0.002)	-0.060^{***} (0.002)
High Fin. Lit.	0.215^{**} (0.087)	0.427^{***} (0.035)	0.327^{***} (0.025)	0.21^{**} (0.009)	-0.009^{***} (0.00)	0.012^{***} (0.00)
Age	0.031^{**} (0.014)	0.407^{***} (0.01)	0.235^{***} (0.005)	0.004 (0.003)	-0.007^{***} (0.000)	-0.006^{***} (0.000)
Age^2	0.000^{**} (0.000)	-0.004^{**} (0.000)	-0.002^{***} (0.000)	0.000^{***} (0.000)	0.000^{**} (0.000)	0.000^{***} (0.000)
$log(\frac{wealth}{income})$	0.928^{***} (0.088)	5.87^{***} (0.088)	2.55^{***} (0.084)	-0.09^{***} (0.019)	-0.184^{***} (0.001)	-0.175^{***} (0.001)
$\frac{wealth}{income}$ Q2	0.613^{***} (0.071)	0.076 (0.075)	1.55^{***} (0.039)	-0.038 (0.029)	0.088^{***} (0.004)	0.033^{***} (0.002)
$\frac{wealth}{income}$ Q3	0.957^{***} (0.184)	-1.293^{***} (0.164)	1.281^{***} (0.010)	-0.134^{***} (0.049)	-0.039^{***} (0.005)	-0.009^{***} (0.004)
$\frac{wealth}{income}$ Q4	0.122 (0.315)	– (–)	1.464^{***} (0.170)	-0.139^{**} (0.067)	-0.117^{***} (0.006)	-0.193^{***} (0.005)
R^2	0.361	0.794	0.541	0.336	0.804	0.730

Notes: Households are divided into three groups according to their self reported financial knowledge: Low (0-4 on scale), intermediate (5-7) and high (8-10). Total wealth is defined by the SCF as the balance between total assets and total debt, and income is the sum of incomes and transfers from all sources. Households are assigned to wealth-to-income quartiles. *** is significant at 1%; ** is significant at 5%; * is significant at the 10% level. Standard errors in the data are computed using the "scfcombo" Stata package in order to account for the SCF complex sample specification as well as the multiple imputation process. Standard errors in the model are computed by simulating $1,000$ bootstrap samples from the SCF data. The wealth-to-income fourth quartile is omitted from the home-ownership logit regression since all simulated households who belong to this quartile end up owning a house.

that is generated by the benchmark model and therefore reduce housing demand elasticities. To what degree does heterogeneity in a certain dimension matter for housing markets is therefore a quantitative question. Our policy experiment suggests that self-assessed financial literacy does play an important role in the housing markets and should hence be incorporated into structural models of housing choice.

Finally, as an aside, we note that the heterogeneous agent model is able to remarkably capture the conditional correlation between between financial literacy and housing market choices, despite targeting only average choices within coarse age groups. We view this evidence as enhancing the validity of the model.

7 Conclusion

We study the role of subjective financial literacy in housing markets. Using SCF data, we document that individuals who self-assess themselves as more financially literate are more likely to own a house and take on more debt against the value of the house. The relationship is economically meaningful and robust to a host of potential confounding factors. Motivated by these empirical patterns, we develop a portfolio choice model with housing to infer the mechanisms that underlie the empirical facts. They key novel feature of the model is that we allow mortgage market parameters and expectations on future house prices to depend on households' subjective literacy.

We estimate the model to match the empirical relationship between subjective financial literacy and housing choices. The estimation reveals that that households with higher subjective financial literacy are in fact more financially savvy - they obtain more attractive mortgage terms and invest in houses that yield higher risk-adjusted returns. Differences in mortgage terms are particularly important for explaining the relationship between literacy and housing choices among young households. Differences in expectations on house price growth are more important for the underlying cross-sectional variation among older households.

A key takeaway of our analysis is that incorporating heterogeneity in financial literacy in an otherwise standard portfolio choice model with housing is quantitatively important for evaluating housing policies. Models that abstract from this heterogeneity overestimate the correlation between homeownership and household wealth and can lead to biased policy evaluations. Our model substantially reduces this bias. More broadly, our results highlight that documenting heterogeneity in financial decision making and incorporating this heterogeneity into structural models is important for understanding the impact of economic policies (Gomes, Haliassos and Ramadorai, 2021).

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan originations and defaults in the mortgage crisis: The role of the middle class." *The Review of Financial Studies*, 29(7): 1635–1670.

Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao. 2017. "Systematic mistakes in the mortgage market and lack of financial sophistication." *Journal of Financial Economics*, 123(1): 42–58.

Agnew, Julie R, and Lisa R Szykman. 2005. "Asset allocation and information overload: The influence of information display, asset choice, and investor experience." *The Journal of Behavioral Finance*, 6(2): 57–70.

Allgood, Sam, and William Walstad. 2013. "Financial literacy and credit card behaviors: A cross-sectional analysis by age."

Armona, Luis, Andreas Fuster, and Basit Zafar. 2019. "Home price expectations and behaviour: Evidence from a randomized information experiment." *The Review of Economic Studies*, 86(4): 1371–1410.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. 2019. "House price beliefs and mortgage leverage choice." *The Review of Economic Studies*, 86(6): 2403–2452.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The economic effects of social networks: Evidence from the housing market." *Journal of Political Economy*, 126(6): 2224–2276.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2022. "Consumer-lending discrimination in the FinTech era." *Journal of Financial Economics*, 143(1): 30–56.

Calvet, Laurent E, John Y Campbell, and Paolo Sodini. 2007. "Down or out: Assessing the welfare costs of household investment mistakes." *Journal of Political Economy*, 115(5): 707–747.

Calvet, Laurent E, John Y Campbell, and Paolo Sodini. 2009. "Measuring the financial sophistication of households." *American Economic Review*, 99(2): 393–98.

Campbell, John Y, and Joao F Cocco. 2003. "Household risk management and optimal mortgage choice." *The Quarterly Journal of Economics*, 118(4): 1449–1494.

Case, Karl E, and Robert J Shiller. 1988. "The behavior of home buyers in boom and post-boom markets."

Case, Karl E, and Robert J Shiller. 1990. "Forecasting prices and excess returns in the housing market." *Real Estate Economics*, 18(3): 253–273.

Charles, Kerwin Kofi, and Erik Hurst. 2002. "The transition to home ownership and the black-white wealth gap." *Review of Economics and Statistics*, 84(2): 281–297.

Cocco, Joao F. 2004. "Portfolio choice in the presence of housing." *The Review of Financial Studies*, 18(2): 535–567.

Cocco, Joao F, Francisco J Gomes, and Pascal J Maenhout. 2005. "Consumption and portfolio choice over the life cycle." *The Review of Financial Studies*, 18(2): 491–533.

Davis, Morris A, Andreas Lehnert, and Robert F Martin. 2008. "The Rent-price ratio for the aggregate stock of owner-occupied housing." *Review of Income and Wealth*, 54(2): 279–284.

Delavande, Adeline, Susann Rohwedder, and Robert J Willis. 2008. "Preparation for retirement, financial literacy and cognitive resources." *Michigan Retirement Research Center Research Paper*, , (2008-190).

Flavin, Marjorie, and Takashi Yamashita. 2002. "Owner-occupied housing and the composition of the household portfolio." *American Economic Review*, 92(1): 345–362.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably unequal? The effects of machine learning on credit markets." *The Journal of Finance*, 77(1): 5–47.

Gargano, Antonio, Marco Giacoletti, and Elvis Jarnecic. 2023. "Local Experiences, Search, and Spillovers in the Housing Market." *The Journal of Finance*, 78(2): 1015–1053.

Gathergood, John, and Jörg Weber. 2017. "Financial literacy: A barrier to home ownership for the young?" *Journal of Urban Economics*, 99: 62–78.

Gaudecker, Hans-Martin Von. 2015. "How does household portfolio diversification vary with financial literacy and financial advice?" *The Journal of Finance*, 70(2): 489–507.

Glaeser, Edward L, and Charles G Nathanson. 2017. "An extrapolative model of house price dynamics." *Journal of Financial Economics*, 126(1): 147–170.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai. 2021. "Household finance." *Journal of Economic Literature*, 59(3): 919–1000.

Greenwald, Daniel L, and Adam Guren. 2024. "Do credit conditions move house prices?" National Bureau of Economic Research.

Guiso, Luigi, Andrea Pozzi, Anton Tsoy, Leonardo Gambacorta, and Paolo Emilio Mistrulli. 2022. "The cost of steering in financial markets: Evidence from the mortgage market." *Journal of Financial Economics*, 143(3): 1209–1226.

Guiso, Luigi, and Tullio Jappelli. 2008. "Financial literacy and portfolio diversification."

Hastings, Justine S, Brigitte C Madrian, and William L Skimmyhorn. 2013. "Financial literacy, financial education, and economic outcomes." *Annu. Rev. Econ.*, 5(1): 347–373.

Jappelli, Tullio, and Mario Padula. 2013. "Investment in financial literacy and saving decisions." *Journal of Banking & Finance*, 37(8): 2779–2792.

Keys, Benjamin J, Devin G Pope, and Jaren C Pope. 2016. "Failure to refinance." *Journal of Financial Economics*, 122(3): 482–499.

Kindermann, Fabian, Julia Le Blanc, Monika Piazzesi, and Martin Schneider. 2021. "Learning about housing cost: Survey evidence from the german house price boom." National Bureau of Economic Research.

Kuchler, Theresa, and Basit Zafar. 2019. "Personal experiences and expectations about aggregate outcomes." *The Journal of Finance*, 74(5): 2491–2542.

Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel. 2023. "Housing market expectations." In *Handbook of Economic Expectations*. 163–191. Elsevier.

Landvoigt, Tim. 2017. "Housing demand during the boom: The role of expectations and credit constraints." *The Review of Financial Studies*, 30(6): 1865–1902.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. 2015. "The housing market (s) of San Diego." *American Economic Review*, 105(4): 1371–1407.

Lusardi, Annamaria, and Olivia S Mitchell. 2007. "Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth." *Journal of monetary Economics*, 54(1): 205–224.

Lusardi, Annamaria, and Olivia S Mitchell. 2011. "Financial literacy and retirement planning in the United States." *Journal of pension economics & finance*, 10(4): 509–525.

Lusardi, Annamaria, and Olivia S Mitchell. 2014. "The economic importance of financial literacy: Theory and evidence." *American Economic Journal: Journal of Economic Literature*, 52(1): 5–44.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S Mitchell. 2017. "Optimal financial knowledge and wealth inequality." *Journal of Political Economy*, 125(2): 431–477.

Ortalo-Magne, Francois, and Sven Rady. 2006. "Housing market dynamics: On the contribution of income shocks and credit constraints." *The Review of Economic Studies*, 73(2): 459–485.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the asymptotics of optimization estimators." *Econometrica: Journal of the Econometric Society*, 1027–1057.

Parker, Andrew M, Wändi Bruine De Bruin, Joanne Yoong, and Robert Willis. 2012. "Inappropriate confidence and retirement planning: Four studies with a national sample." *Journal of behavioral decision making*, 25(4): 382–389.

Piazzesi, Monika, Martin Schneider, and Selale Tuzel. 2007. "Housing, consumption and asset pricing." *Journal of Financial Economics*, 83(3): 531–569.

Shiller, Robert J. 2007. "Understanding recent trends in house prices and home ownership."

Sommer, Kamila, Paul Sullivan, and Randal Verbrugge. 2013. "The equilibrium effect of fundamentals on house prices and rents." *Journal of Monetary Economics*, 60(7): 854–870.

Stanton, Richard, and Nancy Wallace. 1998. "Mortgage choice: What's the point?" *Real estate economics*, 26(2): 173–205.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. 2011. "Financial literacy and stock market participation." *Journal of Financial Economics*, 101(2): 449–472.

Van Rooij, Maarten CJ, Annamaria Lusardi, and Rob JM Alessie. 2012. "Financial literacy, retirement planning and household wealth." *The Economic Journal*, 122(560): 449–478.

Yao, Rui, and Harold H Zhang. 2005. "Optimal consumption and portfolio choices with risky housing and borrowing constraints." *The Review of Financial Studies*, 18(1): 197–239.

Appendix

Contents

A Figures **38**

B Dynamic Programming Solution **41**

B.1 Bellman Equation for $a \geq Ret$ 41

B.2 Bellman Equation for $a = Ret - 1$ 41

B.3 Transformed Model 42

B.3.1 Transformed Model for $a \geq Ret$ 42

B.3.2 Transformed Model for $a < Ret - 1$ 43

B.3.3 Transformed Model for $a = Ret - 1$ 44

C Estimation Procedure **45**

C.1 Standard Errors 46

A Figures

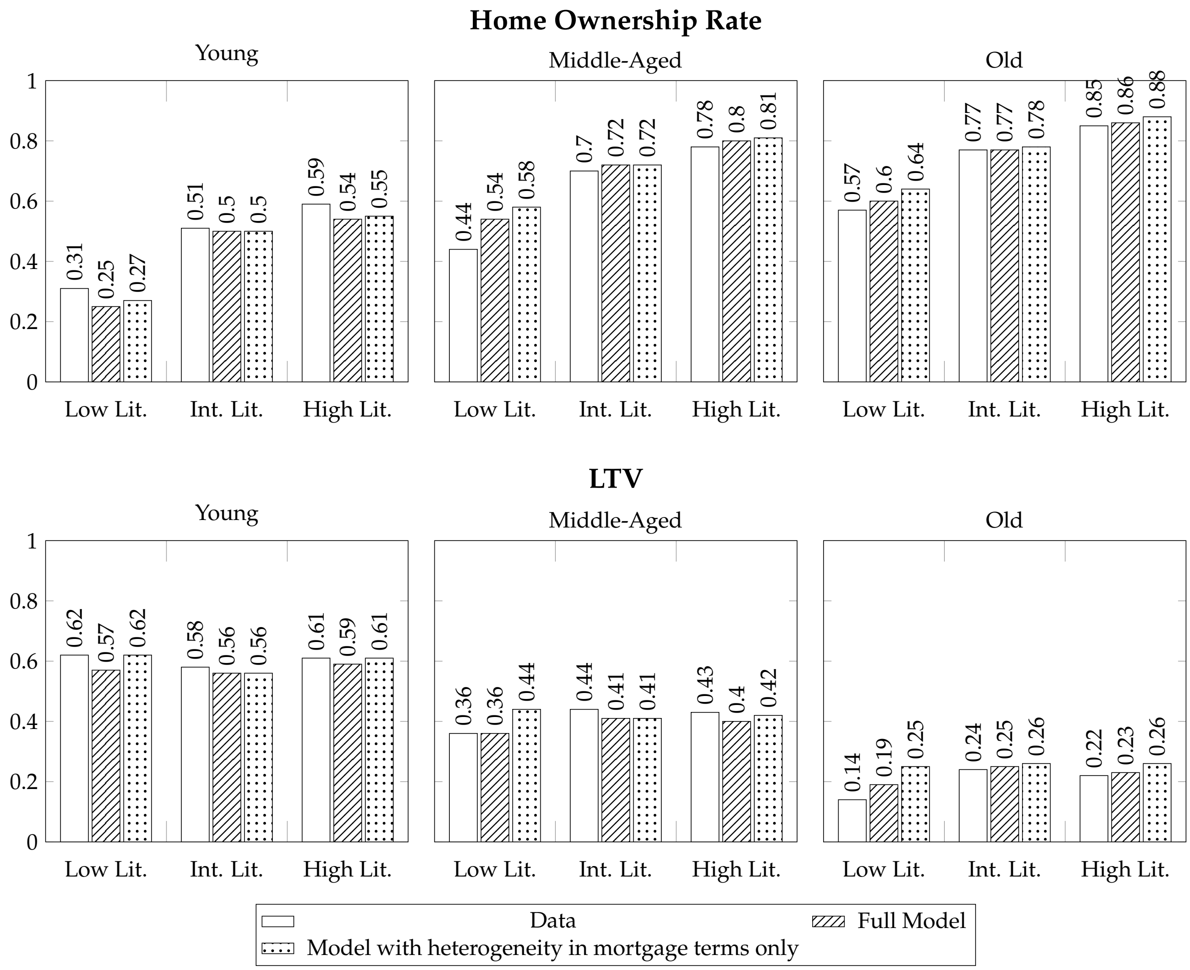

Figure 4: Shutting off Heterogeneity in Expectations on Future Prices

Note: The figure compares between 1) SCF data moments; 2) The full heterogeneous model generated moments ; and 3) Moments generated by a model in which mean expected return ($\mu(f_i)$) and volatility ($\sigma(f_i)$) are set to their benchmark model estimates (Table 5) for all literacy groups $f_i = \{Low, Int, High\}$ whereas estimates of mortgage spread ($\varrho(f_i)$) and down-payment requirements ($\delta(f_i)$) are taken from the full heterogeneous-agent model.

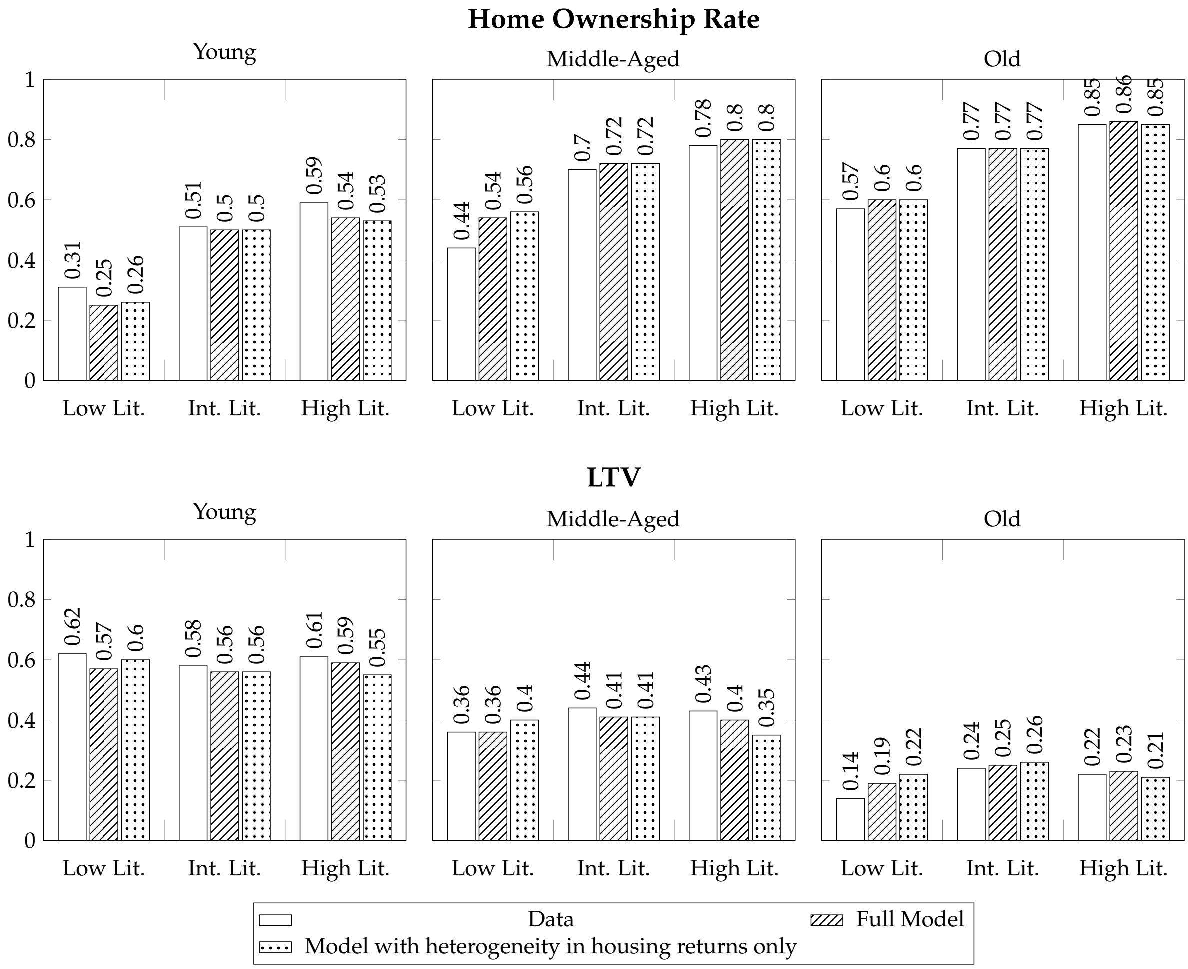

Figure 5: Shutting off Heterogeneity in Mortgage Terms

Note: The figure compares between 1) SCF data moments; 2) The full heterogeneous model generated moments; and 3) Moments generated by a model in which mortgage spread ($\varrho(f_i)$) and down-payment requirements ($\delta(f_i)$) are set to their benchmark model estimates (Table 5) for all literacy groups $f_i = \{Low, Int, High\}$ whereas estimates of mean expected return ($\mu(f_i)$) and volatility ($\sigma(f_i)$) are taken from the full heterogeneous agent model.

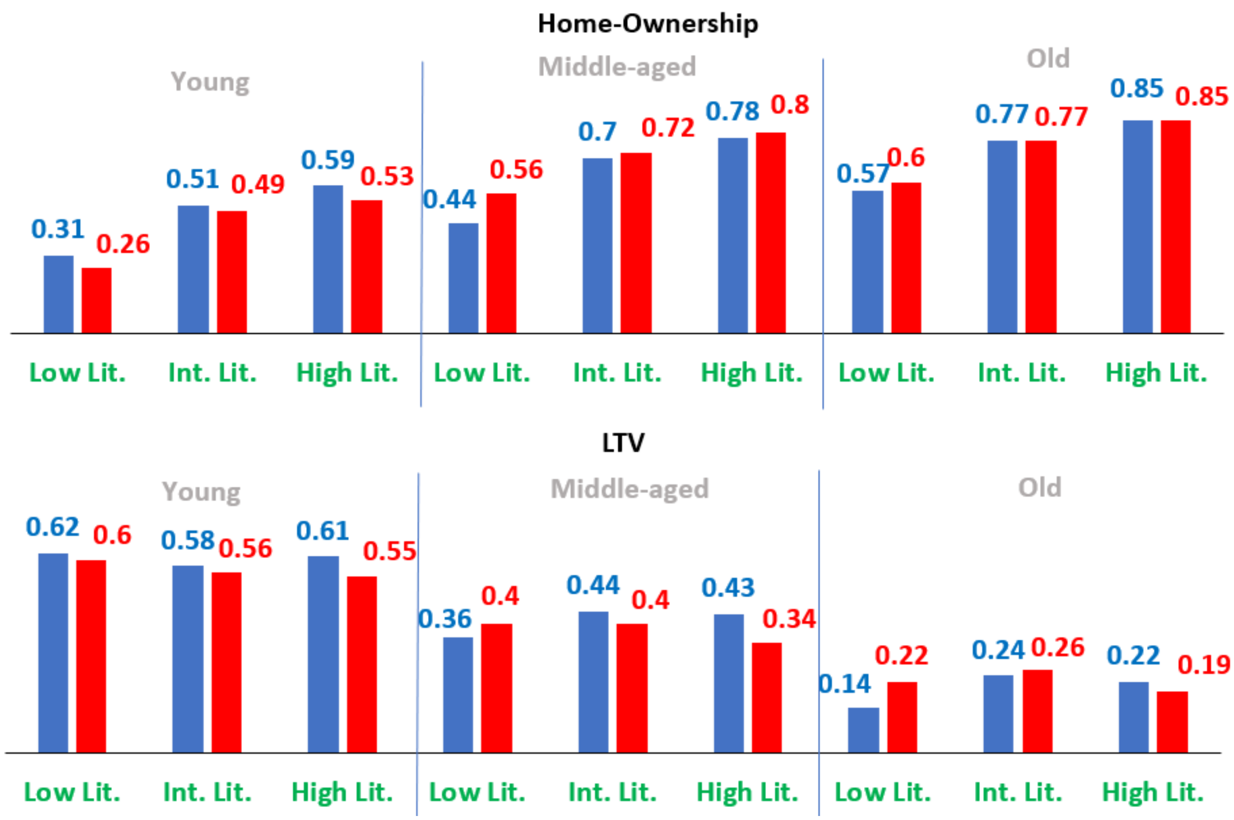

Figure 6: Distorted Expectations

Note: The figure compares between SCF data moments (in blue) and moments generated by a model in which $\mu(f_i), \sigma(f_i), \varrho(f_i), \delta(f_i)$ are set to their baseline model estimates (Table 5) but in which the realized idiosyncratic shock to house prices is drawn from a common distribution $g_{i,t} \sim N(\mu, \sigma)$ where μ and σ are set to their weighted average values from Table 5 (in red).

B Dynamic Programming Solution

Equation 7 specifies the problem faced by household i at age $a < Ret - 1$. For completeness, we will first specify the equivalent problem for $a \geq Ret - 1$. Next, we will present a transformation to the model that serves two purposes. The first is improving on efficiency of computation by reducing the state space. As seen below, we are able to dispose of both the permanent income $\hat{Y}_t$ and the house price index P_t, thereby allowing for enhanced speed in the estimation procedure. Second, the transformed problem is the basis for comparing the model output to the Survey of Consumers Finance survey data.

B.1 Bellman Equation for $a \geq Ret$

Define the state variable tuple $X_t^{Ret} = \{a_t, f_i, W_t, P_t, \tau_{t-1}, (1 + g_{i,t}) H_{t-1}, Y_{Ret}, M_t\}$ where Y_{Ret} is the household's retirement income which is a fraction θ_{Ret} of their income in the period prior to retirement. The following Bellman equation specifies the household value function after retirement, i.e. for $a \geq Ret$:

$$\begin{aligned} &\tilde{V}(a_t, f_i, W_t, P_t, \tau_{t-1}, (1 + g_{i,t}) H_{t-1}, Y_{Ret}, M_t) = \\ &\lambda_{a_t} \Bigg\{ \max_{Z_t} u(C_t, S_t) + \\ &+\beta \mathbb{E}_t^i [\tilde{V}(a_t + 1, f_i, W_{t+1}, P_{t+1}, \tau_t, (1 + g_{i,t+1}) H_t, Y_{Ret}, M_{t+1})] \Bigg\} + \\ &+(1 - \lambda_{a_t}) D(W_t, P_t), \end{aligned} \tag{9}$$

where Z_t is the vector of policy variables defined as $Z_t = \{C_t, H_t, B_t, \tau_t, \xi_t\}$. The problem is subject to the collateral constraint (Equation 3) and budget constraint (Equations 4-6).

B.2 Bellman Equation for $a = Ret - 1$

Next, consider the problem faced by household i one period before retirement, i.e. at age $a = Ret - 1$. Note that in this period the continuation value function is given by $\tilde{V}(.)$, whereas the current value function is given by $V(.)$ (Equation 7). Applying the notation of X_t and X_t^{Ret} previously defined, the household value function at age $a = Ret - 1$ is:

$$
\begin{aligned}
V(X_t) = & \lambda_{a_t} \left\{ \max_{Z_t} u(C_t, S_t) + \beta \mathbb{E}_t^i [\tilde{V}(X_{t+1}^{Ret})] \right\} \\
& + (1 - \lambda_{a_t}) D(W_t, P_t).
\end{aligned} \tag{10}
$$

Note that the state variable in the current value function is the permanent income component at age $a = Ret - 1$, i.e $\hat{Y}_t$, whereas in the continuation value function the state variable is Y_{Ret}. Note that $Y_{Ret} = \theta_{Ret} \exp\left(f(Ret-1) + log\overline{Y_t} + log\hat{Y}_t^i + u_t \right)$ is a state variable at $t+1$ (i.e. part of X_{t+1}^{Ret}). This means that the vector of state variables at age $a = Ret - 1$ includes also $log\overline{Y_t} + u_t$ as a state. The problem is subject to the usual collateral constraint and budget constraint.

B.3 Transformed Model

In order to reduce the state space dimensionality and efficiently compute the policy functions, this section presents a transformed and equivalent household problem. The solution relies on the homothetic nature of the problem and closely follows Landvoigt (2017).

B.3.1 Transformed Model for $a \geq Ret$

By backward induction, consider first the problem defined by equation 9 for households that are retired or are about to retire at the end of the period, i.e. for $a \geq Ret$. We normalize all the quantities of the model by total income Y_{Ret} and use the notation $\tilde{x}$ to denote the normalized variables:

$$
\tilde{w}_t = \frac{W_t}{Y_{Ret}},\ \tilde{ph}_{t-1} = \frac{P_t\,(1+g_{i,t})\,H_{t-1}}{Y_{Ret}}\ \tilde{c}_t = \frac{C_t}{Y_{Ret}}
$$

$$
\tilde{b}_t = \frac{B_t}{Y_{Ret}},\ \tilde{h}_t = \frac{H_t}{Y_{Ret}},\ \tilde{s}_t = \frac{S_t}{Y_{Ret}}.
$$

Denote by $\tilde{v}(a_t, f_i, \tilde{w}_t, \tau_{t-1}, \tilde{ph}_{t-1}, M_t) = \frac{V(a_t, f_i, W_t, P_t, \tau_{t-1}, \left(1+g_{i,t}\right) H_{t-1}, Y_{Ret}, M_t)}{(Y_{Ret} P_t^{-\rho})^{1-\gamma}}$ the normalized value function. Denote the normalized policy variables by $\tilde{z}_t = \left\{\tau_t, \tilde{b}_t, \tilde{h}_t, \tilde{c}_t, \xi_t\right\}$ and the normalized state variables by $\tilde{x}_t = \left\{a_t, f_i, \tilde{w}_t, \tau_{t-1}, \tilde{ph}_{t-1}, M_t\right\}$. Finally, the normalized bequest function is $d(\tilde{w}_t) = \overline{D} \frac{\tilde{w}_t^{1-\gamma}}{1-\gamma}$.

The household problem in 9 can then be re-written as follows:

$$\tilde{v}(\tilde{x}_t) = \lambda_{a_t}\left[\max_{\tilde{z}_t} u(\tilde{c}_t, \tilde{s}_t) + \beta \mathbb{E}_t\left[\tilde{v}(\tilde{x}_{t+1})\left(G^P_{t+1}\right)^{-\rho(1-\gamma)}\right]\right] \ldots$$
$$+ (1-\lambda_{a_t}) d(\tilde{w}_t),$$

where $\tilde{s}_t = \tilde{h}_t$ and $G^P_{t+1} = \frac{P_{t+1}}{P_t} = exp\{\epsilon^P_{t+1}\}$. To recall, $\epsilon^P_{t+1} \sim N(0, \sigma^2_P)$. This problem is subject to a normalized collateral constraint:

$$\tilde{b}_t \geq \begin{cases} 0 & \tau_t = 0 \\ -\left[1-\delta(f_i)\right]\tilde{h}_t & \tau_t = 1 \end{cases}. \tag{11}$$

The problem is also subject to a normalized budget constraint. Specifically, a previous renter faces the following budget constraint:

$$\tilde{c}_t + \tilde{b}_t + \tilde{h}_t\left[(1-\tau_t)\alpha + \tau_t(1+\psi)\right] = \tilde{w}_t,$$

a previous owner who sells faces the following budget constraint:

$$\tilde{c}_t + \tilde{b}_t + \tilde{h}_t\left\{(1-\tau_t)\alpha + \quad \tau_t \quad (1+\psi)\right\} = \tilde{w}_t + (1-\nu)\tilde{ph}_{t-1},$$

and a previous owner who doesn't sell faces the following budget constraint:

$$\tilde{c}_t + \tilde{b}_t + \psi\tilde{ph}_{t-1} = \tilde{w}_t.$$

Note that relative to the original Bellman equation for $a \geq Ret$ (Section B.1), the normalized Bellman equations does not require keeping track of the price P_t and the income at retirement Y_{Ret} as state variables.

B.3.2 Transformed Model for $a < Ret - 1$

Next, consider the case of a household of age $a < Ret - 1$. In this case we normalize quantities by the permanent income $\hat{Y}_t$:

$$w_t = \frac{W_t}{\hat{Y}_t},\; ph_{t-1} = \frac{P_t\left(1+g_{i,t}\right)H_{t-1}}{\hat{Y}_t},\; c_t = \frac{C_t}{\hat{Y}_t}$$

$$b_t = \frac{B_t}{\hat{Y}_t},\; h_t = \frac{H_t}{\hat{Y}_t},\; s_t = \frac{S_t}{\hat{Y}_t}.$$

Denote by $x_t = \{a_t, f_i, w_t, \tau_{t-1}, ph_{t-1}, M_t\}$ the vector of state variables and by

$$v(a_t, f_i, w_t, \tau_{t-1}, ph_{t-1}, M_t) = \frac{V(a_t, f_i, W_t, P_t, \tau_{t-1}, (1+g_{i,t})\, H_{t-1}, \hat{Y}_t, M_t)}{(\hat{Y}_t P_t^{-\rho})^{1-\gamma}}$$

the normalized value function. In addition let $z_t = \{\tau_t, b_t, h_t, c_t, \xi_t\}$ the vector of policy variables in the normalized problem. . Finally, the normalized bequest function is $d(w_t) = \overline{D}\frac{w_t^{1-\gamma}}{1-\gamma}$.

The normalized household problem can then be re-written as follows:

$$\begin{aligned} v(a_t, f_i, w_t, \tau_{t-1}, ph_{t-1}, M_t) = \lambda_{a_t} \Big\{ & \max_{z_t} u(c_t, s_t) ... \\ & + \beta \mathbb{E}_t[(v(a_t+1, f_i, w_{t+1}, \tau_t, h_t, M_{t+1})[G^Y_{t+1}(G^P_{t+1})^{-\rho}]^{1-\gamma}] \Big\} ... \\ & + (1-\lambda_{a_t}) d(w_t), \end{aligned}$$

where $G^Y_{t+1} = \frac{\hat{Y_{t+1}}}{\hat{Y}_t} = \exp\left(\bar{\epsilon}_{t+1} + \hat{\epsilon}^i_{t+1}\right)$ and G^P_{t+1} is defined as above. It is useful to define $\epsilon^{\hat{Y}}_t = \overline{\epsilon_t} + \hat{\epsilon^i_t}$ so that $\epsilon^{\hat{Y}}_t \sim N(0, \sigma^2_{\hat{Y}})$ and $G^Y_t = exp(\epsilon^{\hat{Y}}_t)$. The problem is subject to the normalized collateral constraint defined above (Section B.3.1).

B.3.3 Transformed Model for $a = Ret - 1$

Finally, we normalize the household problem for age $a = Ret - 1$. Using the notation defined above and some algebra, the normalized household problem in this case can be written as:

$$\begin{aligned} v(x_t) = \lambda_{a_t} & \left\{ \max_{z_t} u(c_t, s_t) + \beta \mathbb{E}_t \left[(\tilde{v}(\tilde{x}_{t+1}) \left((G^P_{t+1})^{-\rho} \theta_{Ret} \exp\left(f(a_t) + log\overline{Y_t} + u_t \right) \right)^{1-\gamma} \right] \right\} \\ & + (1-\lambda_{a_t}) d(w_t). \end{aligned}$$

Note that the current value function is $v_t(.)$ while the continuation value function is $\tilde{v}_t(.)$. Also note that in the continuation value, the expression $\theta_{Ret} \exp\left(f(a_t) + log\overline{Y_t} + u_t\right)$ is the factor that allows to convert normalized variables by $\hat{Y}_{Ret-1}$ to variables normalized by Y_{Ret}, e.g. $\tilde{w}_{Ret} = \frac{w_{Ret-1}}{\theta_{Ret} \exp\left(f(Ret-1) + log\overline{Y_{Ret-1}} + u_{Ret-1}\right)}$. The problem is subject to the normalized collateral constraint defined above (Section B.3.1).

C Estimation Procedure

We estimate the model by applying a Simulated Method of Moments (SMM) to the cross-sectional 2016 SCF data. Denote the parameters that we estimate by

$$\eta = \left\{ \left\{ \mu(f_i), \sigma^2(f_i), \delta(f_i), \varrho(f_i) \right\}_{f_i = low, intermediate, high}, \beta, \overline{D}, \sigma_P^2 \right\}.$$

All other parameters are calibrated exogenously and discussed in Section 4.

We begin by drawing a large number of I households from the SCF data. For each sampled household, we denote the vector of sampled state variables by

$$\Omega_t^i = \left\{ a_{i,t}, f_{i,t}, \tau_{i,t-1}, W_{i,t}, P_t(1+g_{i,t})H_{it-1}, \widetilde{Y}_{i,t}, M_{i,t} \right\},$$

where $a_{i,t}$ is the age of the head of the household, f_{it} is the subjective financial literacy category household i belongs to, $\tau_{i,t-1}$ denotes the house ownership status at the beginning of the period, W_{it} is the total wealth, and $P_t(1+g_{i,t})H_{it-1}$ is the house price for owners as defined in Section 4.1. $\widetilde{Y}_{it}$ denotes the household's permanent income in case the household is not retired (i.e. $\widetilde{Y}_{it} = \hat{Y}_{it}$ if $a_{i,t} < Ret$) and denotes the households' income in case the household is retired (i.e. $\widetilde{Y}_{it} = Y_{Ret}$ if $a \geq Ret$). Since the data does not distinguish between the permanent income component $\hat{Y}_{it}$ and the temporary income component, for non-retired households we decompose the observed labor income Y_{it} by simulating the transitory shock from its specified distribution. Similarly, we simulate a moving shock M_{it} for each household based on the calibrated moving probabilities.

Denote the normalized vector of state variables by $\omega_t^i = \{a_{i,t}, f_{i,t}, \tau_{i,t-1}, w_{i,t}, ph_{t-1}, M_{it}\}$, where $w_{it} = \frac{W_{i,t}}{\widetilde{Y}_{it}}$ and $ph_{t-1} = \frac{P_t(1+g_{i,t})H_{it-1}}{\widetilde{Y}_{it}}$. Given the sample of simulated state variables $\omega_t = \{\omega_t^i\}_{i=1}^I$, given the exogenously calibrated parameters, and given a guess for η, we obtain the period t optimal policies for each household i by solving the household problem specified in Section B.3. Denote these policies by $z_t(\omega_t, \eta) = \{\tau_{i,t}, h_{i,t}, c_{i,t}, b_{i,t}, \xi_{i,t}\}_{i=1}^I$ where $h_{i,t} = \frac{H_{i,t}}{\widetilde{Y}_{it}}$, $c_{i,t} = \frac{C_{i,t}}{\widetilde{Y}_{it}}$ and $b_{i,t} = \frac{B_{i,t}}{\widetilde{Y}_{it}}$. We then simulate the period $t+1$ shock to income ($\epsilon_{i,t+1}^{\hat{Y}}$), the shock to price per quality unit of housing (ϵ_{t+1}^P), and the idiosyncratic shock to house price growth ($g_{i,t+1}$). For each household, this allows us to map the policies $z_t(\omega_t^i, \eta)$ into the household's year $t+1$ vector of normalized state variables ω_{t+1}^i. For each households, we convert the normalized state variables at time $t+1$ back to its non-normalized format. That is, we obtain $\Omega_{t+1}^i$for $i = 1, ..., I$.

Next, we compute moments from the simulated $t+1$ sample, i.e. based on $\left\{\Omega_{t+1}^i\right\}_{i=1}^I$. Namely, we define 3 age groups (young, for ages 26 to 40, middle-aged, for ages 41to 60 and old, for ages 61 to 80), and compute the average homeownership rate and the average loan-to-value ratio for homeowners, for each age group and conditional on the subjective financial literacy of the household (low, intermediate, and high). We also compute the average wealth, the average wealth at age 80, and the implied volatility of house prices growth. When computing these moments, we apply the SCF household-specific weights. Denote the vector of these 21 moments by $\bar{\bar{\Theta}}(\eta)$.

Finally, we compute the same moments from the simulated SCF data, i.e. based on $\{\Omega_t^i\}_{i=1}^I$, and denote them by $\bar{\Theta}$. Our estimate for η, denoted by $\hat{\eta}_{SMM}$, is obtained by minimizing the mean of the square error of the simulated moments with respect to their empirical counterpart :

$$\hat{\eta}_{SMM} = \underset{\eta}{argmin} \sum \left(\bar{\Theta} - \bar{\bar{\Theta}}(\eta) \right)^2.$$

C.1 Standard Errors

The standard errors of the estimated SMM parameters are calculated based on Pakes and Pollard (1989). Specifically, we use the fact that $\hat{\eta}_{SMM}$, satisfies :

$$\hat{\eta}_{SMM} = \underset{\eta}{argmin} (\bar{\Theta} - \bar{\bar{\Theta}}(\eta))'(\bar{\Theta} - \bar{\bar{\Theta}}(\eta)).$$

Denote by J the Jacobian matrix of the function $\eta :\longrightarrow \bar{\bar{\Theta}}(\eta)$. Denote by Ω the asymptotic variance of $\bar{\Theta}$. It can be shown that:

$$\sqrt{I}(\hat{\eta}_{SMM} - \eta) \xrightarrow{d} N\left(0, (1 + \frac{1}{s})(J'J)^{-1}J'\Omega J(J'J)^{-1}\right),$$

where s is the number of model simulations. We compute J numerically by calculating small changes of the function $\eta :\longrightarrow \bar{\bar{\Theta}}(\eta)$ at $\eta = \hat{\eta}_{SMM}$. We estimate Ω by bootstrapping the data. We set $s = 100$, i.e. we repeat the SMM estimation 100 times, each time drawing (potentially) different shocks between period t and $t+1$. The asymptotic variance of $\hat{\eta}_{SMM}$, from which we identify the standard errors of $\hat{\eta}_{SMM}$, is given by:

$$Var(\hat{\eta}_{SMM}) = \frac{1}{n}(1 + \frac{1}{s})(J'J)^{-1}J'\hat{\Omega}J(J'J)^{-1}.$$